





GPLCTEHRARVGTAKGIPRISVREPMOTGKAVDSLVLCRCQRELIIGDRQTGKTA  
 VAIDRIHQKRENEAEKKNLYCIYVAIGOKRSTVAOIVKRLHSDALKYTIIVISAT  
 ASDAAPLOYLAPISGANGEPFRNGKHALIIVDDLSKAOAVAYROMSLRLRPPGREA  
 YPDYFYLHSLERAKAMNPDGGSLSLTALPVIEQAGDVSAIPTNIVISITDIOIF  
 LTELVEYKGIPIPAINGVSGSRVSAQVAKMGOVAGSMKLDIGEVREYAAFAQSGD  
 LDASTONLLNRGRTLLKOGQTPMAVEDQAVLFAVGRHLDKIDPAKVTKEFEQ  
 FIAHRSNHOALLDTIRTEGQSPQTEATLKDVLKFLDTFEG  
 178..1695  
 /evidence-not\_experimental  
 /product="mitochondrial ATP synthase alpha subunit  
 precursor"  
 BASE COUNT 635 a 559 c 551 g 610 t 1 others  
 ORIGIN

Query Match 4.98; Score 23; DB 45; Length 2356;  
 Best local Similarity 78.08; Pred. No. 2,94e-01;  
 Matches 32; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Db 1483 gaccacagtcgtgcttccttcgtgctgctgctgctgctgctcatt 1523  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 3 GAACAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 43

RESULT 5 CEC33B4 30112 bp DNA INV 01-OCT-1996  
 LOCUS Caenorhabditis elegans cosmid C33B4.  
 DEFINITION 248367  
 ACCESSION 671818  
 NID  
 KEYWORDS ankyrin; chloride channel protein; FAED.  
 SOURCE Caenorhabditis elegans.  
 ORGANISM Caenorhabditis elegans.

Eukaryotes; mitochondrial eukaryotes; Metazoa; Nematoda;  
 Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea;  
 Rhabditidae; Caenorhabditis.  
 1 (bases 1 to 30112)  
 2 (bases 1 to 30112)

Direct Submission  
 Submitted (11-FEB-1995) Nematode Sequencing Project, Sanger Centre,  
 Hinxton, Cambridge CB10 1HQ, England and Department of Genetics,  
 Washington University, St. Louis, MO 63110, USA. E-mail:  
 jes@sanger.ac.uk or wrenematode.wustl.edu  
 2 (bases 1 to 30112)  
 Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M.,  
 Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,  
 Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Pavello, A.,  
 Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,  
 Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N.,  
 Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,  
 O'Callaghan, M., Parsons, J., Percy, C., Rifkin, L., Roopra, A.,  
 Saunders, D., Showkneen, R., Smalton, N., Smith, A., Sonhammer, E.,  
 Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,  
 Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,  
 Wilkinson-Spratt, J. and Wohlman, P.  
 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*

Nature 368 (6466), 32-38 (1994)  
 94150718  
 Current sequence finishing criteria for the *C. elegans* genome  
 sequencing consortium are that all bases are either sequenced  
 unambiguously on both strands, or on a single strand with both a  
 dye primer and dye terminator reaction, from distinct subclones.  
 Exceptions are indicated by an explicit note.  
 IMPORTANT: This sequence is NOT necessarily the entire insert of  
 clone C33B4. It may be shorter because we only sequence  
 overlapping sections once, or longer because we arrange for a small  
 overlap between neighbouring submissions.  
 The true left end of clone C33B4 is at 1 in this sequence. The true  
 right end of clone C33B4 is at 9858 in  
 sequence CEC05D12.

The true left end of clone C05D12 is at 30012 in this sequence. The  
 true right end of clone B0491 is at 4602 in this sequence. Coding  
 sequences below are predicted from computer analysis, using the  
 program GeneFinder (P. Green, ms in preparation), and other

JOURNAL  
 MEDLINE  
 COMMENT

available information.  
 The start of this sequence (1..100) overlaps with the end of  
 sequence CEB0491.  
 The end of this sequence (30012..30112) overlaps with the start of  
 sequence CEC05D12.  
 Location/Qualifiers  
 1..30112  
 /organism="Caenorhabditis elegans"  
 /clone="C33B4"  
 /chromosome="II"  
 join(<1..43,413..532,582..713,1122..1255,1465..1555,  
 1773..1844,1889..2204,2251..2494,2539..2629,2683..2867,  
 2916..3168,3214..3361,3418..3643,3693..3881,4265..4480)  
 /note="similar to chloride channel protein"  
 /codon\_start=1  
 /product="C33B4.1"  
 /db\_xref="PID:g1066899"  
 /translation="DHVESNETFAEFTCTROAQNYYHFLVEDWFLSALLGITAVLSVG  
 MDVAIEVLOHARVTFYDKMLAISYLAFSQVAHVILVTMEALFCQIVSKOAVGSGI  
 PEVKYIMHGFKNYLTITLVAKVGLTILAMGGGLPIGEGFVEMGAIVATLLSKI  
 TSACQSAFFNEGEMEMLSGCAVGIACITFSAPIGAVLYAIESKTSKFAVANYRG  
 FLAATCSAIFVRCANFVTABQSGTITAFYOTREPTDCFLVEELPFLLLGFLISGLMG  
 SFETIHRQISIFRSKNRYKLIERNFLAFTVFMFAVGVLTFFNGLGIFVAGRLTF  
 RETMADFNNCTWNSNDRCPDITLTHWTGGSEGVSIETSLVLYLYLYFAVLCI  
 SINPAGVFVPSFLIAGAGRLMGETMVLFPPEGMGPGGPPHPGLYAVVKYPLPD  
 AVTHLSVSVICELTQGLSPILPVLIAVMGNACKFLOPISPIVESIRVKYPLPD  
 LPSPSVHTVKVQLMVDVITIKMTYREKELIQLAPHLSPFIRVDHENKILL  
 GSAVKRYLTMLRRHLVYNQDSRNIGMTPESEIFNTIRKTSMLSRSPSRSSNSQ  
 TRENSVPETARSADPSQVEFVSERTISGNTLSISPLHCPNPIPLQAVTRSSADL  
 SKETNADILLNRNIDLEIDAAPFVLGSLKYVHTLFLSLGSHAYVPTDCGLK  
 VGVGLKELRDAMANIYVRGAVPVKPDRLTSGTLFDMSHMKDKLINTPTQTPANG  
 LSNMSPKFGNSLTVPMPAL"  
 complement(join(4885..5105,5542..5761,5808..5895,  
 5947..6127,6408..6603))  
 /note="weak similarity with the E. coli outer membrane  
 protein FAED (Swiss Prot accession number P06970)"  
 /codon\_start=1  
 /product="C33B4.2"  
 /db\_xref="PID:g671820"  
 /translation="MEINOKREEDDEMEWRDGVKRDNSDFSNYCRQRLSSDATI  
 MSQFFVYLIHFVLDIIVARAFFIDRPRGLGLQLPTGMDTDSRGPFGCLYTAGD  
 KAGQKPTVDPSLPGQTASTFGRSAPFTTHMSAVYTEDLVWDAGFAGVANGVNH  
 GLNVRKFDADAPADLSSSDGMYQFELSAASIGSVDFSKREVSGNFNPLIPGVNEL  
 FDFGRFVKGKGGGILNMEFPLTSDPNERAPYTFKLYNMDRHHVGHVVPNV  
 NLFWVKDKTMRMLNRLNPTMVG"  
 complement(join(19454..19683,20307..20460,20598..20745,  
 21230..22071,22309..22631,23139..23411,23469..23810,  
 23857..24445,24516..24617,24877..25146,25254..25313))  
 /note="possesses proline rich domains and has similarity  
 to ankyrin domains; cDNA EST CEES296f comes from this  
 gene"  
 /codon\_start=1  
 /product="C33B4.3"  
 /db\_xref="PID:g671821"  
 /translation="MNOEETVNLQIVPELVNVRKFLAVTQNDIFDVKRKLATLPO  
 ALPOAFNYGLFLPDCDGRAGKFLLEDRTIRDYFTDCVYPLELKYKRYMLNLDKQ  
 OLKAMHYGLKLFCEMDYVQOKNNEKCEKMSQGLDHFDAQGETFUTLAAGIPNRA  
 VYSLIGGAHVDFERNSEGTAMKHAFLSFEENVKTLIELGASPNRPYRIGTPIYI  
 NMTADSNQVAEILLREAADIGVTDHMGHEIHOACKNGLTKVPHLLYFGQIDAE  
 NYNGNSPLHVCVANNRPECARVLLRCADHLAVNKQGTALHVSHTVGNFVADYVQA  
 HNPSSVPIRGTPQYSRRLSSITTRRSMSOSSICQDVYRTPQSVKRGPMNSAAPS  
 PPSKSSRTTIPSEIGTMRSGMDSMRGGMTAAGHETNIARILVPRGVKGFIL  
 RGAKHVAMPLNFEPTAQVLOFFEGVDMSGMAVRAGLRFGDVLLEIDGIDVRCSHD  
 EVVEFIQAQGTITLKVITVDVADMSRGGTIVHRPPTDTHDAGVDYIYAEINRAYS  
 ESRHASVORSGSRNVSAGLNRTTFOAAPTSTFEYNCSRTPOLSRMDSFDFSD  
 SVYDMKRRGQNNVAVSAGLNRTTFOAAPTSTFEYNCSRTPOLSRMDSFDFSD  
 DEDEMPAPASYSIPDLORSSMORSEYSRPRTSRPTKPTPPPPPMQHNQHOY  
 QQQHPSLPRASSTPDIQOQSSSTPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP  
 PLPPPTSSGAPPPPPPPPPGLMHVAASAPVLMNSKGISADALSKVOLKAEPRPESA  
 ASVNNKNNNNNTTDFQMDLKNLAKRKRKRVADHDEDERESRFGSLSRVFNEN  
 VYERKGIQNIQVGNKSGYTSRTSLEPSESEKDRPHFLSDHSPVQVORVLLISQ  
 HLEDNYGQDNMSVASSSTASSSTVLTLPKGFVPSHVIPVDYDDDDPQSGDSD  
 GEIRCSIEFHEHKVDVMSVDDVIGWLSLHSEYTPAFRSORINGRCLRQCRSRET







```

/note="silent"
/replace="c"
1034
/gene="murine tumour necrosis factor receptor 2"
/note="ser to phe"
/replace="t"
1047
/gene="murine tumour necrosis factor receptor 2"
/note="silent"
/replace="t"
1143
/gene="murine tumour necrosis factor receptor 2"
/note="silent"
/replace="t"
1173
/gene="murine tumour necrosis factor receptor 2"
/note="silent"
/replace="a"
1262
/gene="murine tumour necrosis factor receptor 2"
/note="Thr to Cys"
/replace="g"
1317
/gene="murine tumour necrosis factor receptor 2"
/note="silent"
/replace="g"
327 a 417 c 371 g 273 t

Query Match 4.7% Score 22; DB 118; Length 1388;
Best Local Similarity 64.1%; Pred. No. 1.34e+00;
Matches 50; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

Db 108 tcagatgtctgtcgaatgtctctctgccaatattgaaacattctgcaacaagac 167
QY 111 TCAGCTGTTGTGACAAATGCTCTCTGCTACCTACCTAAACACACTGTACAGCAA 170
Db 168 ctggacacgtgtgtgc 185
QY 171 GTGGAAGACCGTGTGCGC 188

RESULT 10
LOCUS MUSTNFR2 1505 bp mRNA ROD 18-APR-1991
DEFINITION Mouse tumor necrosis factor receptor 2 mRNA, complete cds.
ACCESSION M60459
KEYWORDS g199827
SOURCE transmembrane protein; tumor necrosis factor receptor.
ORGANISM Mouse adult macrophage, cDNA to mRNA.
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
Murinae; Mus.

REFERENCE
1 (bases 1 to 1505)
Lewis, M., Tartaglia, L.A., Lee, A.L., Bennett, G.L., Rice, G.C.,
Wong, G.H.W., Chen, E.Y. and Goeddel, D.V.
Cloning and expression of cDNAs for two distinct murine necrosis
factor receptors demonstrate one receptor is species specific
Proc. Natl. Acad. Sci. U.S.A. 88, 2830-2834 (1991)
91187885

FEATURES
Location/Qualifiers
1..1505
/organism="Mus musculus"
41..106
/codon_start=1
41..1465
/codon_start=1
/product="murine tumour necrosis factor receptor 2"
/db_xref="PID:g199828"
/translation="MAPAALWALVFELQWLWATGHTVPAQVLTTPKPEPGYECQISQ
EYDRKACMKCAPPGQYVKHFNKTSIDTVACDCEASMTQVWNOFPTCLSCSSCT
TDQVEIRACTKQONRYCACEAGRYCALKTHSGRCQMRKSCGPGFGVASSRAPNG
VLCKACAPGTFDTSSTDCVCRPHRCSILAPGNASTDAVCAPESTLSAIPRTLTV

BASE COUNT 327 a 417 c 371 g 273 t

Query Match 4.7% Score 22; DB 118; Length 1388;
Best Local Similarity 64.1%; Pred. No. 1.34e+00;
Matches 50; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

Db 108 tcagatgtctgtcgaatgtctctctgccaatattgaaacattctgcaacaagac 167
QY 111 TCAGCTGTTGTGACAAATGCTCTCTGCTACCTACCTAAACACACTGTACAGCAA 170
Db 168 ctggacacgtgtgtgc 185
QY 171 GTGGAAGACCGTGTGCGC 188

RESULT 10
LOCUS MUSTNFR2 1505 bp mRNA ROD 18-APR-1991
DEFINITION Mouse tumor necrosis factor receptor 2 mRNA, complete cds.
ACCESSION M60459
KEYWORDS g199827
SOURCE transmembrane protein; tumor necrosis factor receptor.
ORGANISM Mouse adult macrophage, cDNA to mRNA.
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
Murinae; Mus.

REFERENCE
1 (bases 1 to 1505)
Lewis, M., Tartaglia, L.A., Lee, A.L., Bennett, G.L., Rice, G.C.,
Wong, G.H.W., Chen, E.Y. and Goeddel, D.V.
Cloning and expression of cDNAs for two distinct murine necrosis
factor receptors demonstrate one receptor is species specific
Proc. Natl. Acad. Sci. U.S.A. 88, 2830-2834 (1991)
91187885

FEATURES
Location/Qualifiers
1..1505
/organism="Mus musculus"
41..106
/codon_start=1
41..1465
/codon_start=1
/product="murine tumour necrosis factor receptor 2"
/db_xref="PID:g199828"
/translation="MAPAALWALVFELQWLWATGHTVPAQVLTTPKPEPGYECQISQ
EYDRKACMKCAPPGQYVKHFNKTSIDTVACDCEASMTQVWNOFPTCLSCSSCT
TDQVEIRACTKQONRYCACEAGRYCALKTHSGRCQMRKSCGPGFGVASSRAPNG
VLCKACAPGTFDTSSTDCVCRPHRCSILAPGNASTDAVCAPESTLSAIPRTLTV

```

```

SOPEPTRSQPLDQEPSPQTSILTSLSGTPILBQSTKGGISLPILGIVGTSLGLIM
LGLVNCILIVQRKKPKSLQDRADKVPVHPVDEKSDQAVGLEQOHLHTTAPSSSSSLES
SASGDRRAPPGHPPQARVMAEAOQFOEARASSRISDSHSGHGTHTVNTCIYNVCS
SDHSSQCSQASATVGDPAKPSAPKDEQVFFSQECPQSQCETITLQSHKPLP
LGPVDMGMPKPSQAGWFDQIAVKA"
BASE COUNT 347 a 459 c 408 g 291 t
ORIGIN

Query Match 4.7% Score 22; DB 89; Length 1505;
Best Local Similarity 64.1%; Pred. No. 1.34e+00;
Matches 50; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

Db 193 tcagatgtctgtcgaatgtctctctgccaatattgaaacattctgcaacaagac 252
QY 111 TCAGCTGTTGTGACAAATGCTCTCTGCTACCTACCTAAACACACTGTACAGCAA 170
Db 253 ctggacacgtgtgtgc 270
QY 171 GTGGAAGACCGTGTGCGC 188

RESULT 11
LOCUS MUSTNFR1 3796 bp mRNA ROD 01-AUG-1991
DEFINITION Murine tumor necrosis factor I receptor (TNFR-1) mRNA, complete
cds.
ACCESSION M59378
NID g202094
KEYWORDS tumor necrosis factor receptor.
SOURCE Mus musculus lymphoid cDNA to mRNA.
ORGANISM Mus musculus
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
Murinae; Mus.

REFERENCE
1 (bases 1 to 3796)
Goodwin, R.G., Anderson, D.M., Jerzy, R., Davis, T., Brannan, C.I.,
Copeland, N.G., Jenkins, N.A. and Smith, C.A.
Molecular cloning and expression of the type 1 and type 2 murine
receptors for tumor necrosis factor
Mol. Cell. Biol. 11, 3020-3026 (1991)
91246168

FEATURES
Location/Qualifiers
1..3796
/organism="Mus musculus"
/cell_line="7B9"
/cell_type="helper T-cell"
/tissue_type="lymphoid"
/map="Chromosome 4"
43..108
/gene="TNFR-1"
/codon_start=1
43..1467
/gene="TNFR-1"
/codon_start=1
/product="tumor necrosis factor receptor 1"
/db_xref="PID:g202095"
/translation="MAPAALWALVFELQWLWATGHTVPAQVLTTPKPEPGYECQISQ
EYDRKACMKCAPPGQYVKHFNKTSIDTVACDCEASMTQVWNOFPTCLSCSSCT
TDQVEIRACTKQONRYCACEAGRYCALKTHSGRCQMRKSCGPGFGVASSRAPNG
VLCKACAPGTFDTSSTDCVCRPHRCSILAPGNASTDAVCAPESTLSAIPRTLTV
SOPEPTRSQPLDQEPSPQTSILTSLSGTPILBQSTKGGISLPILGIVGTSLGLIM
LGLVNCILIVQRKKPKSLQDRADKVPVHPVDEKSDQAVGLEQOHLHTTAPSSSSSLES
SASGDRRAPPGHPPQARVMAEAOQFOEARASSRISDSHSGHGTHTVNTCIYNVCS
SDHSSQCSQASATVGDPAKPSAPKDEQVFFSQECPQSQCETITLQSHKPLP
LGPVDMGMPKPSQAGWFDQIAVKA"
109..1464
/gene="TNFR-1"
/codon_start=1
/product="tumor necrosis factor receptor 1"
/db_xref="PID:g202095"
/translation="MAPAALWALVFELQWLWATGHTVPAQVLTTPKPEPGYECQISQ
EYDRKACMKCAPPGQYVKHFNKTSIDTVACDCEASMTQVWNOFPTCLSCSSCT
TDQVEIRACTKQONRYCACEAGRYCALKTHSGRCQMRKSCGPGFGVASSRAPNG
VLCKACAPGTFDTSSTDCVCRPHRCSILAPGNASTDAVCAPESTLSAIPRTLTV

BASE COUNT 863 a 1052 c 981 g 900 t
ORIGIN Chromosome 4.

Query Match 4.7% Score 22; DB 90; Length 3796;
Best Local Similarity 64.1%; Pred. No. 1.34e+00;

```



EDBSAKAFADAYIREKRESVASGDDRAEDMDEALEKGEAEQSEFEADREDAEDAR  
 EEEPEKMEADYVMAVVDKAAEAGAEQYGLTPTKQLGQAQSGRGPFASSHDE  
 TLPGSESEATADENRDEQPEETATSGTQSTIELSSPTPMDENRDMVDE  
 TNEEESQSEFVATYKESLYSQEISKPAVTPPLNGFSEGKTDATDKDNASA  
 SNLPSSENEEKFRSALRDYCEVKAATLIDKDSISAVSEKSPSKSPSLPS  
 PSLPKTLPGKRSVNFSLTNEIKVSAEAVTADISQVVEEHCAPDEKTLV  
 VPSQSTGAGHTIYQSPDKEKSHLPTVEIKPAPVPSPEFSAKDENERASVS  
 PMDEVPDSEPIEIVLSPKPLIGESAYESFLSADDKASGRGAESPENKSKQ  
 GSPDVPVSEMTSLVDQKQEGKSTDFPIKEDFGQEKTDDEVAMSQPALADE  
 RKLGVSPQIDVSGFSGKEDTKMSIQEYVSKSAIPVDEGVAEDTYSHMEGVASV  
 STASVATSPFPTDDVPSLHAEVGSFSTEDVDSLSVSVVOTPTTFOETEMPSK  
 EECPRMISPPDFSPKAKSKTPQDHRSEQSSMSIEFQSPESGLANDMERQSPD  
 HFTVAGVLIHTEGPTDYSDMDQSSLSHKIPMEPEPTQNDLSIELTSVOV  
 EASPTSSAHTPSQDSTLSDVAPPRMSLYLIPMEPEPTQNDLSIELTSVOV  
 SPLTPRESSPLTSTDSATKAKTACHSSSPPIIDAASEKVLGKSPKDI  
 HLALNRDLSPGLEKSGGKPGDVSAYQKPEETRSPEDEYDIESEYKTRISD  
 VGGYIEKIERITKSPSGYSYETICKTKPEDGYSYETIEKTRIPEGGYSYD  
 ISEKTPSPVSGYSYETKERSRLDDIANGIDDDSDGHTLGDPSYETIEKITS  
 PPSGYSYETKTRIPDSTICYETAKTRTPQASTYSETSDLCYTAEKKSPS  
 EARQDVLCVSSCEYKHPKTELSPFINPNLEWFASEPTESKPLTQSGGAPP  
 PGKQKQDCDEPTTSPSESAPQSDSDVPPETECPSITADANDESEETPTD  
 KVTYKHMDPPAPQDPSRPHDPVMDVPEALAEONLKGALKDKLEKTKKP  
 GTKTKSSPVKSKGKPLAASPAGLAKESDDKSVRAKPKKESVKAAPKPTTP  
 EVKAARBEKDKETKNAANASAKSATATAGTGTKTKSSAVPPGLVYLDLCYIP  
 NHSKNVDVFEKFRVSSYVSGNDPAAEPFRAVLALLEGKAQGSNMQVTLIP  
 THDSEVMEWQETHEKQQLNIMVLASSSTVVMQDESPACKIEL"

BASE COUNT 2912 a 2238 c 2212 g 2054 t

Query Match 4.7%; Score 22; DB 83; Length 9416;  
 Best Local Similarity 73.9%; Pred. No. 1.34e+00;  
 Matches 34; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Db 1514 GCAAGAGCTGACGTACGTCAGCAGGAGTGCATCGCACCCACAA 1559  
 QY 260 GCAAGAGCTGACGTACGTCAGCAGGAGTGCATCGCACCCACAA 305

RESULT 14 RCUS0550 13356 bp DNA VRL 31-JAN-1997  
 LOCUS Rat cytomegalovirus major DNA binding protein, ICP18.5,  
 DEFINITION glycoprotein B and DNA polymerase genes, complete cds.  
 ACCESSION U0550  
 NID 91808955  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Rat cytomegalovirus.  
 Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 Betaherpesvirinae; Muromegalovirus.  
 1 (bases 1 to 13356)  
 Beuken, E., Slobbe, R., Bruggeman, C.A. and Vink, C.  
 Cloning and sequence analysis of the genes encoding DNA polymerase,  
 glycoprotein B, ICP 18.5 and major DNA-binding protein of rat  
 cytomegalovirus  
 J. Gen. Virol. 77 (Pt 7), 1559-1562 (1996)  
 96335691  
 2 (bases 1 to 13356)  
 Beuken, E., Slobbe, R., Bruggeman, C.A. and Vink, C.  
 Direct Submission  
 Submitted (05-MAR-1996) Cornelis Vink, Medical Microbiology,  
 University of Limburg, PO Box 5800, Maastricht 6222 WH, The  
 Netherlands  
 Location/Qualifiers  
 1..13356  
 /organism="Rat cytomegalovirus"  
 /strain="Maastricht"  
 319..4164  
 /note="similar to HCMV Ad169 UL57, major DNA binding  
 protein, encoded by GenBank Accession Number X17403"  
 /codon\_start=1  
 /product="major DNA binding protein"  
 /db\_xref="PID:g1255111"  
 /translation="MAEDDLANLAPVAAAALWLVRRDRAGEVLSVLSICDRECPVV

VSPLLDLTVDFEVGTVTRTPMSCDDGVLTVRTSECPVFVLYFGTDEVLVSVDHGD  
 VRLCESSRRFRGVPAPRDRDPTDVAALCRLRLHDPERTLGIYACGNGKLEMLYA  
 GOLVPCLEAVSVQIARSREGKVPFLPATLIPDVGGRGDESGADSSGSGSDEE  
 FVKGHLVPALECLIFVFTSWGQCMRVDTARLIDAGRQFVEDSOLKAKLAPFK  
 RYHGSQKLSJLERDQMLVMDVSAELAFSSVYFDSYFPGGAFDESWPVVYKSA  
 VSGHELVRDLRDLMLHSLHVGALFFANNVSVQTRIGLSAAGGGGGGGGGLGGG  
 LGGHGGGGGGGGGAGGNGVATFRESLLRSVRFVGLTAMYEALDSKRPVFG  
 TEADDCYCHLAFYCGTQPLVSSVMVFNEMGVYSGVSGDHAYTQHVHVSFNG  
 CEACGGRCCTCATPFRMHTEWQIPKAVKEPTVATLVARPADVGLNGFRRE  
 GPERDNGDAARSEDGKAGVPGSGVAGAHGVPATSLPASGCGGPGAGGGG  
 GPTLDRMKHLNVOYCKNSLIDSTDEDFVARSREFYSVIGNLRVIDESVFR  
 LAEYRAAGGDEIAGCTOFFENLDLNPFTQSFPLLYAYYRVVYVLAIONLNAS  
 GYVNDPLTPQVSRNQHOFVYAGAFATPLKKGFLHRLDTKILKSEPELLMDFR  
 AFASACVYKASIRKACMSVRSVLSACIKNRKAVCSLRGPAATFYRHAQRKPLIQ  
 GYAFLLYCHDRMFNCEVTCLEFNHVRFQNSLPDSYDGRDEEDALVFLAATN  
 EYENDVETTPDPCIANVYEFHNFELFGLRDLYSTIQALGRVTVQVNRHCPYL  
 TSGMPFGSAEYVLEFKMKKLEGPAPAKAAVRESALRTVDFRSLTYSFLEKF  
 TCGMGTDRIDFQGOVATYVSGVRSLSITSAGAQDFRGLRSLLRVLDLVGRS  
 WRDSYVDSQVRSRVLAALDGDPPDVEATVIAEIMAGRDGEVPEVDVDFVD  
 QEHVALLREKLIALVARGVDFSAANREVLTRGSGTSSSTGRRRPPVLRPVLGRG  
 RVLRRRRRPLQQQRRRAVRPRRGGGGGPGVRAGRDTPDSAVISSRGEADDAPLP  
 PAQRRRGRDLQPKGRSK"

4438..7119

/note="similar to HCMV Ad169 UL56 encoded by GenBank  
 Accession Number X17403"  
 /codon\_start=1  
 /product="ICP18.5"  
 /db\_xref="PID:g1255112"  
 /translation="MNTLQKLCVCKNECAMDAECLKYCDPSIVLMDSTAFORNGA  
 TWIHLRYTLPALVAQNTAOTSVLSYLEMLAGLITMRELDTALDFAHRDRERY  
 YRIINLDGCPSPVTLFAPGLHLIELATLIERLCKINCYVGLDRTQGVVEGR  
 RLLSLGALCDPSVAGTRETSTCYCEELMVAVNOGRSINRIEGLCDHVTV  
 KVTYQDNDAGTVEQDLGEVAIRAPVKSIIRSLKNLASSGASCAZYINDAEALRG  
 YNLFSEIFRYSLSDTYTSKTSIVIRVHVVTMRQNLNASHGLMKALRGLSRFLYG  
 EAVEDIFSLGEGFAGHERMFVGSISAPKVDMTSMVSKSPENNLNRLHESNE  
 YAKRSLSEIRDAADPGTAATGTSSTGGTSGTSGTSGTSGTSGTSGTSGTSGT  
 PERGRSSSHGTSRPTSRAGGAGRRGADTSAEGSGAAGARADARADKCGGV  
 GYTRGDPPEAAASTTAPARMNFALCEGDPDLLRTHDVAKENVNRKARADLVKVS  
 YNKNVACIKSOEHLVTLVNLVGTVCLEALSKVMNGFLSRQATKEARALPDVGS  
 LAYDHLAVVNLVHKLPLVELLPHLGQVYTFINGFLTRYLDRLPYLVNMYAC  
 DNGGLIYHVKEDLVRCADGTAEPDGMVAGYQGFSEFVDVNDVNDIQKAVHARVLI  
 LSVLNETFGKLDVRVEDDGAAGDGIYLVNREAPLLRYAGRRFMSKLDYLLLY  
 KHLALCDEASGHGVPGPVSLVAPRAGPPADATANRRDPSPSGGRKRKRTS  
 LVDLVADADRLADGEVVPDSMRVHPAPVAPGCGDR"

7025..9769

/note="similar to HCMV Ad169 UL55, glycoprotein B, encoded  
 by GenBank Accession Number X17403"  
 /codon\_start=1  
 /product="glycoprotein B"  
 /db\_xref="PID:g1255113"  
 /translation="MRTAWPTARWCPILCACILSLPCFVPVGVTAESSTTPAPETT  
 PARAGPTAKAAASVGHTHHHHGHSHSTVEPASGPTTYSPPYDSDQIPSVSTG  
 KPYRICMALSTDLVRFKTECPAPSKPTPEEGIMLVTKQDIVAITFKVITYKNL  
 OFORSYAYIWDVILGTVNRLAMPCEIDQINRNSCYSTQAVRMDYVYIAHRDEE  
 ENVTMLIPDDYSSSSRSTVEDTAYHRRKVMYTESCTINCVVTVTKARSNRPS  
 FVLSSGEVETPFFNGTDEEDTANRVEANYEMVDRFGDWNVTHRVHMAF  
 LERPEVTAWEIQKSEALCEKRWQVVKRAIRTAONSYSYHVSRLTATVTCGKH  
 HNINQIVNCLNGTEELDRVTEETHYKVDGNVEVYOTSGGLVFWQIKPKR  
 LHELGEAINITENVTSEGVNNTSRSRDRDVSVDIKNDIYTLQDTYLOFTYLDLR  
 YINGALONILEANCLDQKQAEMLRELKINFSNLSIYSPVTAAGVALLSECE  
 VPVDDSVKILKDMRIFVNDKVNVCARPHVLFKFNVSNSKIESGQLEHNLNHR  
 TERCEPSRKLIIDEVEGFEDYVFNKVNLTQIDELDTMIGVLESLEPLEDFOILE  
 LYSRGEIRASNFNLEETKREYNQKOSVRLFTQIIEPTPLRGLDFOGGLGAA  
 KGIMGVAGAVAGSIAGSFVFTNFTVTIIIVIAVLVIYIFRRQQAQVNR  
 FVEYFFPHAMQVAGKPLGAGDGRPPSYEDSTSQGRKNVTKGDDGSHKGGSG  
 KAYSGEDAAAMLALQDLEESGRQAAGVAKVKNRILDRAGNSGYLSLPSDVEDR

9779..13141

/note="similar to HCMV Ad169 UL54, DNA polymerase, encoded  
 by GenBank Accession Number X17403."  
 /codon\_start=1  
 /product="DNA polymerase"  
 /db\_xref="PID:g1255114"

e-mail: [rw@nematode.wustl.edu](mailto:rw@nematode.wustl.edu) and [jes@sanger.ac.uk](mailto:jes@sanger.ac.uk)

. AC

5,

## CONCLUSION

rSDI:

5

SSS  
GFT

**YTND**

.220

22113...22238,22286...22483,22536...22762,23207...23460,  
23510...23802,24333...24563,24612...24966,25012...25154,  
25836...25940,26403...26437,28217...28299))  
/gene="K05B2.5"  
/codon\_start=1  
/evidence=not\_experimental  
/db\_xref="PID:g868197"  
/translation="MILQKLVLPCLKIKSYTFRRIIDRKOTTAPHWHHLSARESNF  
EEDQDQIGSLNDQLTNYSSOYINMGVOQVEPLVPOAPDGGYGVIVFAAFMSNLV  
VDGISTAFSEFKQSKERYQESDGLTVFIGSLIIGTYLLVGLVSLCNCNKIEQRIVVM  
AGLISGIAFMIAFASPNIYVFMILYGVNMGGLFGMIIPALIVVGVFFDSCRAMATG  
ISVAGSGVTFLLPLICQVCIANLGWQKTIWLVAYMIFSCVFYGYLYRPLPMDLNEI  
RQEMEPKQALSKVEDDEDAVEPHNPRTSAASGAIDEAVSQNPQVORLSILT  
EGDEGETPKDAHKVRSHTATRSKHTMTSNSSQHDCLKSRGNLSSDNRLSRVSAR  
DYNOSLSKLSKSGGASNLISAMSGVDPKFEFQPMNRQDIFYGGGIONLKEFKEGTWA  
GYRASTLSIPRSVGVQAVSOLNLSRTGSLGGPGIAEDEEMISPEFVDDGCKVILPL  
MNAELSEMDLNLKNTMMILCISNLLGMGFYIPIMFLKDSMKLMSLAWFLV  
PIFVNTIGRYVEFGWLTDOKILSALTINNLISLVSGLLTLACPLLTISIAGQYFVAV  
FGFIICKLFPYQHLEFAAYICLTISIVLADLMGLENLTNSFGLLVARGIASLVGPPF  
AGIVDITGSYDTAFYFGLVLLVAGLISATIPFVTKKADATADGMPQLQDQDNVSGK  
LSVLTERSEALTQYRTIQSMKQHQHLLQDLEEKQKQKQGVVEEGNEMDDNVCS  
NNDNDKLMFY"  
BASE COUNT 12511 a 6305 c 6467 g 12316 t  
ORIGIN

Query Match 4.7%; Score 22; DB 37; Length 37599;  
Best Local Similarity 84.4%; Pred. No. 1.34e+00;  
Matches 27; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
Db 6692 aaaaataatagaatcaaggaagtag 6723  
|||||  
QY 420 AAAATTAATTAGGATCATGCAAGTCAGATAG 451

Search completed: Tue Dec 2 17:04:37 1997  
Job time : 437 secs.

\*\*\*\*\*

WAPREH

(TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

MPsrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm  
Run on: Tue Dec 2 16:51:33 1997; MasPar time 315.17 Seconds  
949.537 Million cell updates/sec  
Circular output not generated.

Title: >US-08-915-004-10  
Description: (1-1089) from US08915004.seq  
Perfect Score: 1089  
N.A. Sequence: 1 ATGAACAAGTGTGCTGCTG.....TAAATAAAGCTGCTTATAA 1089  
Comp: TACTGTGTCACGACGAC.....ATTATTTCGACGATATT

Scoring table: TABLE default  
Gap 6  
Nmatch STD : Dbase 0; Query 0  
Searched: 359085 seqs, 137405154 bases x 2  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: EST-STS-THREE  
1:EST199 2:EST200 3:EST201 4:EST202 5:EST203 6:EST204  
7:EST205 8:EST206 9:EST207 10:EST208 11:EST209 12:EST210  
13:EST211 14:EST212 15:EST213 16:EST214 17:EST215  
18:EST216 19:EST217 20:EST218 21:EST219 22:EST220  
23:EST221 24:EST222 25:EST223 26:EST224 27:EST225  
28:EST226 29:EST227 30:EST228 31:EST229 32:EST230  
33:EST231 34:EST232 35:EST233 36:EST234 37:EST235 38:EST236  
39:EST237 40:EST238 41:EST239 42:EST240 43:EST241 44:EST242 45:EST243  
46:EST244 47:EST245 48:EST246 49:EST247 50:EST248 51:EST249 52:EST250  
53:EST251 54:EST252 55:EST253 56:EST254 57:EST255 58:EST256 59:EST257 60:EST258 61:EST259 62:EST260 63:EST261 64:EST262 65:EST263 66:EST264 67:EST265 68:EST266 69:EST267 70:EST268 71:EST269 72:EST270 73:EST271 74:EST272 75:EST273 76:EST274 77:EST275 78:EST276 79:EST277 80:EST278 81:EST279 82:EST280 83:EST281 84:EST282 85:EST283 86:EST284 87:EST285 88:EST286 89:EST287 90:EST288 91:EST289 92:EST290 93:EST291 94:EST292 95:EST293 96:EST294 97:EST295 98:EST296 99:EST297 100:EST298 101:EST299 102:EST300 103:EST301 104:EST302 105:EST303 106:EST304 107:EST305 108:EST306

Database: EST-STS-FOUR  
49:gnEST1 50:gnEST2 51:gnEST3 52:gnEST4 53:gnEST5  
54:gnEST6 55:gnEST7 56:gnEST8 57:gnEST9 58:gnEST10  
59:gnEST11 60:gnEST12 61:gnEST13 62:gnEST14 63:gnEST15  
64:gnEST16 65:gnEST17 66:gnEST18 67:gnEST19 68:gnEST20  
69:gnEST21 70:gnEST22 71:gnEST23 72:gnEST24 73:gnEST25  
74:gnEST26 75:gnEST27 76:gnEST28 77:gnEST29 78:gnEST30 79:gnEST31 80:gnEST32 81:gnEST33 82:gnEST34 83:gnEST35 84:gnEST36 85:gnEST37 86:gnEST38 87:gnEST39 88:gnEST40 89:gnEST41 90:gnEST42 91:gnEST43 92:gnEST44 93:gnEST45 94:gnEST46 95:gnEST47 96:gnEST48 97:gnEST49 98:gnEST50 99:gnEST51 100:gnEST52 101:gnEST53 102:gnEST54 103:gnEST55 104:gnEST56 105:gnEST57 106:gnEST58 107:gnEST59 108:gnEST60

Statistics: Mean 11.077; Variance 1.866; scale 5.938  
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result Query  
SUMMARIES

No.	Score	Match	Length	DB	ID	Description	Pred. No.
1	220	20.2	463	65	AA233719	zr47g08.r1 Soares NhH	0.00e+00
2	220	20.2	463	83	HS1151638	zr47g08.r1 Soares NhH	0.00e+00
3	74	6.8	530	16	AA195113	zr35a03.r1 Soares NhH	6.00e-105
4	74	6.8	530	54	AA195113	zr35a03.r1 Soares NhH	6.00e-105
5	22	2.0	400	39	G10922	human STS SHGC-13782	9.42e-05
6	22	2.0	472	77	AT4774	5740 Arabidopsis thal	9.42e-05
7	21	1.9	245	40	G11923	human STS MR4116	1.84e-03
8	21	1.9	253	38	G06067	human STS WI-6406	1.84e-03
9	21	1.9	377	90	MM05110	mb89g02.r1 Soares mou	1.84e-03
10	21	1.9	392	53	T03759	IB862 Infant brain, B	1.84e-03
11	21	1.9	431	24	AA219045	zq16e09.sl Stratagene	1.84e-03
12	21	1.9	443	60	AA213094	mw82b02.r1 Soares mou	1.84e-03
13	21	1.9	443	22	AA213094	mw82b02.r1 Soares mou	1.84e-03
14	21	1.9	443	98	MMAA13451	21953 Arabidopsis tha	1.84e-03
15	21	1.9	496	76	AT55319	21953 Arabidopsis tha	1.84e-03
16	21	1.9	589	22	AA210214	mu40g03.r1 Soares 2Nb	1.84e-03
17	21	1.9	589	62	AA210214	mu40g03.r1 Soares 2Nb	1.84e-03
18	21	1.9	589	97	MMAA12293	mu40g03.r1 Soares 2Nb	1.84e-03
19	21	1.9	1035	95	MM54110	mb68f10.r1 Soares mou	1.84e-03
20	20	1.8	204	106	HSC5200	Human chromosome 5 LA	3.11e-02
21	20	1.8	204	74	HUMC5200	Human chromosome 5 LA	3.11e-02
22	20	1.8	204	46	HUMC5200	Human chromosome 5 LA	3.11e-02
23	20	1.8	267	43	G25882	human STS EST49741	3.11e-02
24	20	1.8	288	104	MMAA904	mu06e03.r1 Soares mou	3.11e-02
25	20	1.8	288	18	AA200904	mu06e03.r1 Soares mou	3.11e-02
26	20	1.8	288	59	AA200904	mu06e03.r1 Soares mou	3.11e-02
27	20	1.8	295	101	MMAA60758	va02g10.r1 Soares mou	3.11e-02
28	20	1.8	345	36	G00944	fruit fly STS Dm0983	3.11e-02
29	20	1.8	348	44	G27630	human STS SHGC-32817	3.11e-02
30	20	1.8	389	8	AA126855	zn88f12.sl Stratagene	3.11e-02
31	20	1.8	400	40	G14220	human STS SHGC-7175 c	3.11e-02
32	20	1.8	400	40	G13445	human STS SHGC-10698	3.11e-02
33	20	1.8	417	63	AA228090	zt55g11.sl Soares NhH	3.11e-02
34	20	1.8	421	83	HS1153832	zt30e09.sl Soares ova	3.11e-02
35	20	1.8	439	97	MMAA11122	mp58f11.r1 Soares 2Nb	3.11e-02
36	20	1.8	439	58	AA111122	mp58f11.r1 Soares 2Nb	3.11e-02
37	20	1.8	439	3	AA111122	mp58f11.r1 Soares 2Nb	3.11e-02
38	20	1.8	448	51	AA155033	ms14c04.r1 Stratagene	3.11e-02
39	20	1.8	469	103	MMAA78725	mt16h02.r1 Soares mou	3.11e-02
40	20	1.8	469	59	AA178725	mt16h02.r1 Soares mou	3.11e-02
41	20	1.8	496	76	AT28118	20321 Arabidopsis tha	3.11e-02
42	20	1.8	504	15	AA189692	mt90c04.r1 Soares mou	3.11e-02
43	20	1.8	510	75	AT18221	21638 Arabidopsis tha	3.11e-02
44	20	1.8	548	7	AA126432	zk94h02.sl Soares pre	3.11e-02
45	20	1.8	2517	38	G06714	human STS WI-7816	3.11e-02

ALIGNMENTS

RESULT 1 AA233719 463 bp mRNA EST 28-FEB-1997  
LOCUS zr47g08.r1 Soares NHMPu Sl Homo sapiens CDNA clone 666590 5'  
DEFINITION AA233719  
ACCESSION G1856711  
NID 91856711  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;  
Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 463)  
REFERENCE Hillier, L., Clark, N., Dubouque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and  
Wilson, R.  
TITLE The WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT

Contact: Wilson RK  
WashU-Merck EST Project  
Washington University School of Medicine



4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.edu

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 448.

Location/Qualifiers

FEATURES

source

1..463

/organism="Homo sapiens"

/note="Organ: mixed (see below); Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2NBHM, pregnant uterus NBHPU, and fetal heart NBH19W) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."

/clone="666590"

/clone.lib="Soares NhMpu S1"

/tissue.type="Pooled human melanocyte, fetal heart, and pregnant uterus"

/lab\_host="DH10B"

<1..>463

154 a 104 c 88 g 117 t

mRNA

BASE COUNT

ORIGIN

Query Match 20.2%; Score 220; DB 65; Length 463;  
Best Local Similarity 99.5%; Pred. No. 0.00e+00;  
Matches 221; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 36 ctcaagtttttggcggaataaaatggcgacacacaccttgaaggccctaaatgcacgca 95  
QY 868 CTCAGTTTGTGCGGATATAAAATGGCGACCAAGACACCTTGAAGGCCCTAATGCAGCA 927

Db 96 ctaagactcaagacgtacacactttcccaaaactgtcactcagagctctaaagaagacc 155  
QY 928 CTAAGAGCACTCAAGACGTACACCTTTCCTCCAAACTGTCTACTAGAGCTCAAGAGACC 987

Db 156 atcaggttccttcacagcttcacaaatgtatcagaagttatttttagaaatg 215  
QY 988 ATCAGGTTCCCTTCAGAGCTTCACAAATGTACAAATGTATCAGAAGTTATTTTAGAAATG 1047

216 atagtaaccagggtccaatcagtaaaataagctgcttataa 257  
QY 1048 ATAGGTAACCAAGTCCATCAGTAAATAAGCTGCTTATAA 1089

RESULT 2  
ID HS1151638 standard; RNA; EST; 463 BP.  
AC AA233719;  
NI 91856711  
DT 06-MAR-1997 (Rel. 51, Created)  
DE 06-MAR-1997 (Rel. 51, Last updated, version 1)  
DT z47908.r1 Soares NhMpu S1 Homo sapiens cDNA clone 666590 5'.  
KW EST.  
OS Homo sapiens (human)  
OC Eukaryotae; mitochondria eukaryotes; Metazoa; Chordata;  
OC Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP 1-463  
RA Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M.,  
RA Holman M., Hultman M., Kucaba T., Le M., Lennon G., Marra M.,  
RA Parsons J., Rifkin L., Rohlfing T., Soares M., Tan F.,  
RA Trevasaki E., Waterston R., Williamson A., Wohlmann P., Wilson R.;  
RT "The WashU-Merck EST Project";  
RL Unpublished  
CC Contact: Wilson RK WashU-Merck EST Project Washington University  
CC School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis,

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email:

est@wustl.edu This clone is available royalty-free through

LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for

further information. Seq primer: -28ml3 rev2 ET from Amersham High

quality sequence stop: 448.

Location/Qualifiers

1..463

/organism="Homo sapiens"

/note="Organ: mixed (see below); Vector: pT73D-Pac

(Pharmacia) with a modified polylinker; Site 1: Not I;

Site 2: Eco RI; Equal amounts of plasmid DNA from three

normalized libraries (melanocyte 2NBHM, pregnant uterus

NBHPU, and fetal heart NBH19W) were mixed, and ss circles

were made in vitro. Following HAP purification, this DNA

was used as tracer in a subtractive hybridization reaction

reaction. The driver was PCR-amplified cDNAs from pools of

5,000 clones made from the same 3 libraries. The pools

consisted of I.M.A.G.E. clones 260232-265223,

340488-345479, and 484488-489479."

/clone="666590"

/clone.lib="Soares NhMpu S1"

/tissue.type="Pooled human melanocyte, fetal heart, and

pregnant uterus"

/lab\_host="DH10B"

<1..>463

Sequence 463 BP; 154 A; 104 C; 88 G; 117 T; 0 other;

mRNA

Query Match 20.2%; Score 220; DB 83; Length 463;

Best Local Similarity 99.5%; Pred. No. 0.00e+00;

Matches 221; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 36 ctcaagtttttggcggaataaaatggcgacacacaccttgaaggccctaaatgcacgca 95

QY 868 CTCAGTTTGTGCGGATATAAAATGGCGACCAAGACACCTTGAAGGCCCTAATGCAGCA 927

Db 96 ctaagactcaagacgtacacactttcccaaaactgtcactcagagctctaaagaagacc 155

QY 928 CTAAGAGCACTCAAGACGTACACCTTTCCTCCAAACTGTCTACTAGAGCTCAAGAGACC 987

Db 156 atcaggttccttcacagcttcacaaatgtatcagaagttatttttagaaatg 215

QY 988 ATCAGGTTCCCTTCAGAGCTTCACAAATGTACAAATGTATCAGAAGTTATTTTAGAAATG 1047

216 atagtaaccagggtccaatcagtaaaataagctgcttataa 257

QY 1048 ATAGGTAACCAAGTCCATCAGTAAATAAGCTGCTTATAA 1089

RESULT 3

LOCUS AA195113 530 bp mRNA EST 17-JAN-1997

DEFINITION zr35a03.r1 Soares NhMpu S1 Homo sapiens cDNA clone 665356 5'.

ACCESSION AA195113

NID 91784803

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryotae; mitochondria eukaryotes; Metazoa; Chordata;

Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 530)

AUTHORS Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M.,

Holman M., Hultman M., Kucaba T., Le M., Lennon G., Marra M.,

Parsons J., Rifkin L., Rohlfing T., Soares M., Tan F.,

Trevasaki E., Waterston R., Williamson A., Wohlmann P., and

Wilson R.

TITLE The WashU-Merck EST Project

JOURNAL Unpublished (1995)

COMMENT

Contact: Wilson RK

WashU-Merck EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (infoimage.llnl.gov) for further information.  
High quality sequence stop: 257.

## FEATURES

source

Location/Qualifiers

1..530

/organism="Homo sapiens"

/note="Organ: mixed (see below); Vector: pT7T3D-Pac  
(Pharmacia) with a modified polylinker; Site.1: Not I;  
Site.2: Eco RI; Equal amounts of plasmid DNA from three  
normalized libraries (melanocyte 2NBHM, pregnant uterus  
NBHPU, and fetal heart NBHH19W) were mixed, and ss circles  
were made in vitro. Following HAP purification, this DNA  
was used as tracer in a subtractive hybridization  
reaction. The driver was PCR-amplified cDNAs from pools of  
5,000 clones made from the same 3 libraries. The pools  
consisted of I.M.A.G.E. clones 260232-265223,  
340488-345479, and 484488-489479."

/clone="665356"

/tissue\_type="Pooled human melanocyte, fetal heart, and

pregnant uterus"

/lab\_host="DH10B"

&lt;1..&gt;530

170 a 96 c 93 g 168 t 3 others

mrna

BASE COUNT

ORIGIN

Query Match 6.8%; Score 74; DB 16; Length 530;  
Best Local Similarity 100.0%; Pred. No. 6.00e-105;  
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 acaaatgtatcagaagtatttttagaataatgtagtgaaccaggtcccaatcagtaaaaa 60  
|||||  
Qy 1016 ACAAAATTGTATCAGAAGTTATTTTAGAAATGATAGGTACCCAGGTCCATCAGTAAAAA 1075  
|||||

Db 61 taagctgcttataa 74

Qy 1076 TAAGCTGCTTATAA 1089

## RESULT

LOCUS

DEFINITION

ACCESSION

NID

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

JOURNAL

COMMENT

TITLE

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

(Pharmacia) with a modified polylinker; Site.1: Not I;  
Site.2: Eco RI; Equal amounts of plasmid DNA from three  
normalized libraries (melanocyte 2NBHM, pregnant uterus  
NBHPU, and fetal heart NBHH19W) were mixed, and ss circles  
were made in vitro. Following HAP purification, this DNA  
was used as tracer in a subtractive hybridization  
reaction. The driver was PCR-amplified cDNAs from pools of  
5,000 clones made from the same 3 libraries. The pools  
consisted of I.M.A.G.E. clones 260232-265223,  
340488-345479, and 484488-489479."

/clone="665356"

/tissue\_type="Pooled human melanocyte, fetal heart, and

pregnant uterus"

/lab\_host="DH10B"

&lt;1..&gt;530

170 a 96 c 93 g 168 t 3 others

mrna

BASE COUNT

ORIGIN

Query Match 6.8%; Score 74; DB 54; Length 530;  
Best Local Similarity 100.0%; Pred. No. 6.00e-105;  
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 acaaatgtatcagaagtatttttagaataatgtagtgaaccaggtcccaatcagtaaaaa 60  
|||||  
Qy 1016 ACAAAATTGTATCAGAAGTTATTTTAGAAATGATAGGTACCCAGGTCCATCAGTAAAAA 1075  
|||||

Db 61 taagctgcttataa 74

Qy 1076 TAAGCTGCTTATAA 1089

## RESULT

LOCUS

DEFINITION

ACCESSION

NID

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

JOURNAL

COMMENT

TITLE

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

5 G10922 400 bp DNA STS 06-OCT-1995  
human STS SHGC-13782 clone pg-6721.  
G10922  
G988028

STS sequence; primer; sequence tagged site.  
human plasmid clones, generated from a lymphoblastoid cell line  
from a human male. Localized to human chromosome 12 by analysis on  
the NIGMS Human/Rodent Somatic Cell Hybrid Panel #1, Coriell  
Institute for Medical Research, Camden, NJ 08103.

Homo sapiens

Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;

Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Choanata;

Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates;

Catarrhini; Hominoidea; Homo.

1 (bases 1 to 400)

Myers, R.M.

Unpublished (1995)

Contact: Richard M. Myers

Stanford Human Genome Center (SHGC)

Stanford University School of Medicine

Department of Genetics, M-344, Stanford, CA 94305, USA

Tel: 4157259687

Fax: 4157259689

Email: myers@shgc.stanford.edu

Primer A: GTACCTCAAGAAAGACCC

Primer B: TGGTGGGAGTATCAGTTC

STS size: 98

PCR Profile:

Initial incubation: 94 degrees C for 90 seconds

Denaturation: 94 degrees C for 15 seconds

Annealing: 62 degrees C for 23 seconds

Polymerization: 72 degrees C for 30 seconds

PCR Cycles: 30

Thermal Cycler: Perkin Elmer 9600

Protocol:

Template: 25 ng

Primer: each 1 uM







constructed and normalized by Bento Soares and M.Fatima

Bonafido.  
/clone="677163"  
/clone\_lib="Soares mouse NML"  
/tissue\_type="Liver"  
/lab\_host="DH10B"  
123 a 105 c 118 g 97 t

BASE COUNT  
ORIGIN

Query Match 1.9%; Score 21; DB 60; Length 443;  
Best Local Similarity 88.9%; Pred. No. 1.84e-03;  
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 223 aaatggagcccaagacacaccttgaagg 249

QY 887 AAATGGCGACCAAGACACCTTGAAGG 913

RESULT 13  
AA213094 443 bp mRNA EST 31-JAN-1997  
mw82b02.r1 Soares mouse NML Mus musculus cDNA clone 677163 5'.

AA213094  
g1811731

KEYWORDS

SOURCE

ORGANISM

Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;

Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Mus

1 (bases 1 to 443)

Marra M., Hillier L., Allen M., Bowles M., Dietrich N., Dubuque T.,

Geisel S., Kucaba T., Lacy M., Le M., Martin J., Morris M.,

Schellenberg K., Steptoe M., Tan F., Underwood K., Moore B.,

Theising B., Wylie T., Lennon G., Soares B., Wilson R. and

Waterston R.

The WashU-HHMI Mouse EST Project

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:416867

Seq primer: -28m13 rev2 ET from Amersham.

Location/Qualifiers

1..443

/organism="Mus musculus"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer [5];

TGTTACCAATCTGAAGTGGAGCGCGCGAATCTTTTTTTTTTTT 3';

Gdouble-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not

I and Eco RI sites of the modified pT73 vector. Library

constructed and normalized by Bento Soares and M.Fatima

Bonafido.

/clone="677163"

/clone\_lib="Soares mouse NML"

/lab\_host="DH10B"

<1..>443

123 a 105 c 118 g 97 t

BASE COUNT

ORIGIN

Query Match 1.9%; Score 21; DB 22; Length 443;

Best Local Similarity 88.9%; Pred. No. 1.84e-03;

Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 223 aaatggagcccaagacacaccttgaagg 249

QY 887 AAATGGCGACCAAGACACCTTGAAGG 913

RESULT 14

ID MAA13451 standard; RNA; EST; 443 BP.

AC AA213094;

NI g1811731

DT 03-FEB-1997 (Rel. 50, Created)

DT 20-FEB-1997 (Rel. 51, Last updated, Version 2)

DE mw82b02.r1 Soares mouse NML Mus musculus cDNA clone 677163 5'.

KW EST.

OS Mus musculus (house mouse)

OC Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;

OC Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

OC Mus.

RN [1]

RP 1-443

RA Marra M., Hillier L., Allen M., Bowles M., Dietrich N.,

Dubuque T., Geisel S., Kucaba T., Lacy M., Le M., Martin J.,

RA Morris M., Schellenberg K., Steptoe M., Tan F., Underwood K.,

RA Moore B., Theising B., Wylie T., Lennon G., Soares B., Wilson R.,

RA Waterston R.

RT "The WashU-HHMI Mouse EST Project";

RL Unpublished.

CC Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project

CC Washington University School of Medicine 4444 Forest Park Parkway,

CC Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810

CC Email: mouseest@wustl.edu This clone is available

CC royalty-free through LLNL; contact the IMAGE Consortium

CC (info@image.llnl.gov) for further information. MGI:416867 Seq

CC primer: -28m13 rev2 ET from Amersham.

CC Location/Qualifiers

Key

source

1..443

/organism="Mus musculus"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer [5];

TGTTACCAATCTGAAGTGGAGCGCGCGAATCTTTTTTTTTTTT 3';

double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT73 vector. Library

constructed and normalized by Bento Soares and M.Fatima

Bonafido.

/clone="677163"

/clone\_lib="Soares mouse NML"

/tissue\_type="Liver"

/lab\_host="DH10B"

<1..>443

mRNA

Seq

Sequence 443 BP; 123 A; 105 C; 118 G; 97 T; 0 other;

SQ

Query Match 1.9%; Score 21; DB 98; Length 443;

Best Local Similarity 88.9%; Pred. No. 1.84e-03;

Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 223 aaatggagcccaagacacaccttgaagg 249

QY 887 AAATGGCGACCAAGACACCTTGAAGG 913

RESULT 15

ID AT35319 standard; RNA; EST; 496 BP.

AC N96353;

NI g1269126

DT 19-APR-1996 (Rel. 47, Created)

DT 12-MAR-1997 (Rel. 51, Last updated, Version 9)

DE 21953 Arabidopsis thaliana cDNA clone G86477.

KW EST.

OS Arabidopsis thaliana (thale cress)

OC Eukaryotae; mitochondrial eukaryotes; Viridiplantae;

OC Charophyta/Embryophyta group; Embryophyta; Magnoliophyta;

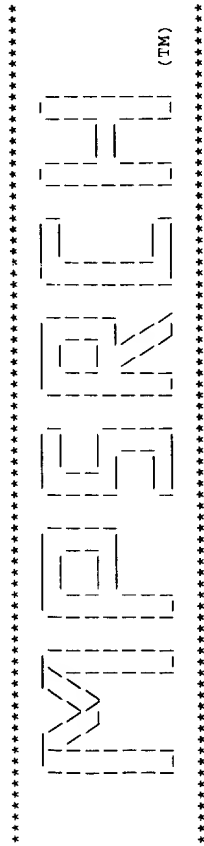
OC Magnoliopsida; Capparales; Brassicaceae; Arabidopsis.  
 RN [1]  
 RP 1-496  
 RX MEDLINE; 95148729.  
 RA Newman T., deBruijn F.J., Green P., Keegstra K., Kende H.,  
 RA McIntosh L., Ohlrogge J., Raikhel N., Somerville S., Thomasow M.,  
 RA Retzel E., Somerville C.;  
 RT "Genes galore: a summary of methods for accessing results from  
 RT large-scale partial sequencing of anonymous Arabidopsis cDNA  
 RT clones";  
 RL Plant Physiol. 106:1241-1255(1994).  
 CC Contact: Thomas Newman MSU-DOE Plant Research Laboratory Michigan  
 CC State University MSU-DOE-PRL Michigan State University, Plant  
 CC Biology Bldg., E. Lansing, MI Tel: 517-353-0834 Fax: 517-353-9168  
 CC Email: 22313tcm@bm.cl.msu.edu. NCBI gi: 1269126  
 FH Key Location/Qualifiers  
 FH  
 FT source 1..496  
 FT /organism="Arabidopsis thaliana"  
 FT /clone="G8G47"  
 FT /strain="var columbia"  
 FT /note="thale cress"  
 FT <1..>496  
 SQ Sequence 496 BP; 132 A; 87 C; 124 G; 137 T; 16 other;

Query Match 1.9%; Score 21; DB 76; Length 496;  
 Best Local Similarity 80.0%; Pred. No. 1.84e-03;  
 Matches 28; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 160 ataacttcgctgctgggaagtcgtttcttatgag 194  
 Cp 769 ATAACCTCAGCAGCTGGAAGTCTGTCTGTGAG 735

Search completed: Tue Dec 2 16:56:58 1997  
 Job time : 325 secs.





\*\*\*\*\*  
Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.  
\*\*\*\*\*  
MPSrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm  
Run on: Tue Dec 2 16:40:50 1997; MasPar time 608.81 seconds  
Circular output not generated.  
Title: >US-08-915-004-10  
Description: - (1-1089) from US08915004.seq  
Perfect Score: 1089  
N.A. Sequence: 1 ATGACACAGTGTCTGTCTGCTG.....TAAATATAGCTCTATATA 1089  
Comp: TACTGTTCACACACACACACAC.....ATTTCATTCGACGAAATTT

Scoring table: TABLE default  
Gap 6  
Nmatch STD : Dbase 0; Query 0  
Searched: 707517 seqs, 256659390 bases x 2  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: EST-STS  
1:EST1 2:EST2 3:EST3 4:EST4 5:EST5 6:EST6 7:EST7 8:EST8  
9:EST9 10:EST10 11:EST11 12:EST12 13:EST13 14:EST14  
15:EST15 16:EST16 17:EST17 18:EST18 19:EST19 20:EST20  
21:EST21 22:EST22 23:EST23 24:EST24 25:EST25 26:EST26  
27:EST27 28:EST28 29:EST29 30:EST30 31:EST31 32:EST32  
33:EST33 34:EST34 35:EST35 36:EST36 37:EST37 38:EST38  
39:EST39 40:EST40 41:EST41 42:EST42 43:EST43 44:EST44  
45:EST45 46:EST46 47:EST47 48:EST48 49:EST49 50:EST50  
51:EST51 52:EST52 53:EST53 54:EST54 55:EST55 56:EST56  
57:EST57 58:EST58 59:EST59 60:EST60 61:EST61 62:EST62  
63:EST63 64:EST64 65:EST65 66:EST66 67:EST67 68:EST68  
69:EST69 70:EST70 71:EST71 72:EST72 73:EST73 74:EST74  
75:EST75 76:EST76 77:EST77 78:EST78 79:EST79 80:EST80  
81:EST81 82:EST82 83:EST83 84:EST84 85:EST85 86:EST86  
87:EST87 88:EST88 89:EST89 90:EST90 91:EST91 92:EST92  
93:EST93 94:EST94 95:EST95 96:EST96 97:EST97 98:EST98  
99:EST99  
EST-STS-TWO  
100:EST100 101:EST101 102:EST102 103:EST103 104:EST104  
105:EST105 106:EST106 107:EST107 108:EST108 109:EST109  
110:EST110 111:EST111 112:EST112 113:EST113 114:EST114  
115:EST115 116:EST116 117:EST117 118:EST118 119:EST119  
120:EST120 121:EST121 122:EST122 123:EST123 124:EST124  
125:EST125 126:EST126 127:EST127 128:EST128 129:EST129  
130:EST130 131:EST131 132:EST132 133:EST133 134:EST134  
135:EST135 136:EST136 137:EST137 138:EST138 139:EST139  
140:EST140 141:EST141 142:EST142 143:EST143 144:EST144  
145:EST145 146:EST146 147:EST147 148:EST148 149:EST149  
150:EST150 151:EST151 152:EST152 153:EST153 154:EST154  
155:EST155 156:EST156 157:EST157 158:EST158 159:EST159  
160:EST160 161:EST161 162:EST162 163:EST163 164:EST164  
165:EST165 166:EST166 167:EST167 168:EST168 169:EST169  
170:EST170 171:EST171 172:EST172 173:EST173 174:EST174

Statistics: Mean 11.043; Variance 1.850; scale 5.968

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description	Pred. No.
1	394	36.2	602	195	AA037313	zc52h03.r1 Soares sen	0.00e-00
2	101	9.3	346	73	H88769	ym23g12.r1 Homo sapie	1.35e-166
3	23	2.1	344	61	H14106	ym62a05.r1 Homo sapie	6.39e-06
4	22	2.0	300	183	AA100384	zn46h08.r1 Stratagene	1.45e-04
5	22	2.0	330	13	RICC0437A	Rice cDNA, partial se	1.45e-04
6	22	2.0	360	164	CI4856	Human fetal brain cDN	1.45e-04
7	22	2.0	385	180	AA087288	mollid06.r1 Life Tech	1.45e-04
8	22	2.0	410	6	T74804	yc60d04.r1 Homo sapie	1.45e-04
9	22	2.0	453	111	N21157	yd72e10.r1 Homo sapie	1.45e-04
10	22	2.0	465	6	T77351	yd72e10.r1 Homo sapie	1.45e-04
11	22	2.0	472	57	T42477	5740 Arabidopsis thal	1.45e-04
12	22	2.0	493	55	RICS15559A	Rice cDNA, partial se	1.45e-04
13	22	2.0	796	117	W29026	55c10 Human retina CD	1.45e-04
14	21	1.9	253	129	HSC0BE062	H. sapiens partial CD	2.90e-03
15	21	1.9	265	9	T89287	yd37f05.sl Homo sapie	2.90e-03
16	21	1.9	291	48	HUM157E01B	Human fetal brain cDN	2.90e-03
17	21	1.9	325	112	HTS3517	A. thaliana transcrib	2.90e-03
18	21	1.9	325	65	H29337	ym60h07.sl Homo sapie	2.90e-03
19	21	1.9	325	19	T54964	yb42d03.r1 Homo sapie	2.90e-03
20	21	1.9	339	88	H65000	yu66d10.sl Homo sapie	2.90e-03
21	21	1.9	371	99	N61165	TgESTzy27b03.r1 Toxop	2.90e-03
22	21	1.9	376	5	T71088	yc50d04.r1 Homo sapie	2.90e-03
23	21	1.9	377	150	W18051	mb83g02.r1 Soares mou	2.90e-03
24	21	1.9	381	164	C14139	Human fetal brain cDN	2.90e-03
25	21	1.9	383	47	H45707	yp23h05.sl Homo sapie	2.90e-03
26	21	1.9	388	76	HUM122D10A	Human fetal brain cDN	2.90e-03
27	21	1.9	392	146	T03759	IB862 Infant brain, B	2.90e-03
28	21	1.9	395	5	T71079	yc50c04.r1 Homo sapie	2.90e-03
29	21	1.9	417	185	AA134912	zo26d02.sl Stratagene	2.90e-03
30	21	1.9	418	157	AA030678	mi22c03.r1 Soares mou	2.90e-03
31	21	1.9	427	157	AA031758	zk14f03.r1 Soares pre	2.90e-03
32	21	1.9	443	198	N43104	SW31CA9398K Brugia ma	2.90e-03
33	21	1.9	447	184	AA011021	ze34c01.sl Soares ret	2.90e-03
34	21	1.9	447	156	AA021559	ze69b06.sl Soares ret	2.90e-03
35	21	1.9	450	131	N68886	TgESTzy39d10.r1 Toxop	2.90e-03
36	21	1.9	453	193	AA164719	zo93f04.sl Stratagene	2.90e-03
37	21	1.9	471	128	W95775	ze07d08.r1 Soares fet	2.90e-03
38	21	1.9	474	62	H17326	ym38e07.sl Homo sapie	2.90e-03
39	21	1.9	478	172	AA061066	mj86h11.r1 Soares mou	2.90e-03
40	21	1.9	484	126	W75508	me55g11.r1 Soares tha	2.90e-03
41	21	1.9	496	134	N96353	21953 Arabidopsis th	2.90e-03
42	21	1.9	499	183	AA100749	z191h06.r1 Stratagene	2.90e-03
43	21	1.9	500	94	N38845	yy80d11.sl Homo sapie	2.90e-03
44	21	1.9	514	178	AA075143	zm86f04.sl Stratagene	2.90e-03
45	21	1.9	551	197	H91502	SWMPCA0865K Brugia ma	2.90e-03

ALIGNMENTS

RESULT 1  
LOCUS AA037313 602 bp mRNA EST 25-NOV-1986  
DEFINITION zc52h03.r1 Soares senescent fibroblasts NbHSF Homo sapiens cDNA  
clone 325973 5'  
ACCESSION AA037313  
NID 91512420  
KEYWORDS EST.

SOURCE ORGANISM	human.	Homo sapiens
REFERENCE	Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 602)	
AUTHORS	Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.	
TITLE	The WashU-Merck EST Project	
JOURNAL	Unpublished (1995)	
COMMENT	Contact: Wilson RK WashU-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert Length: 1203 Std Error: 0.00 Seq primer: -28M13 rev2 from Amersham High quality sequence stop: 435. Location/Qualifiers 1..602 /organism="Homo sapiens" /note="Vector: pT73D (Pharmacia) with a modified polylinker V.TYPE: Phagemid; Site_1: Not I; Site_2: Eco RI; TGTACATTCGAATGGAGCGCGCAATTTTTTTTTTTTTTTT 3', double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo." /clone="325973" /tissue_type="senescent fibroblast" /lab_host="DH10B (ampicillin resistant)" <1..>602	
BASE COUNT	209 a 128 c 136 g 126 t	3 others
ORIGIN		
Query Match	36.2%; Score 394; DB 195; Length 602;	
Best Local	Similarity 99.5%; Pred. No. 0.00e+00;	
atches	406; Conservative 0; Mismatches 0; Indels 2; Gaps 2;	
Db	2 tctaaagcaccctgtagaaaacacacaaaattcgagtgcttttggtctctgctaactca 61	
QY	466 TCTAAAGCACCCCTGTAGAAACACACAAAT-GCAGTGTCTTGGTCTCCTGCTAACTCA 524	
Db	62 gaaaggaaatgcaacacacacacacacatatgttcgcgaaacagtgaaatcaactcaaaatg 121	
QY	525 GAAAGGAAATGCACACACACACACATATGTTCCGGAACAGTGAACTCAACTCAAAATG 584	
Db	122 tggaaatgatgttaccctgtgtgaggaggcattcttcaggtttgctgttcctacaaagt 181	
QY	585 TGGAAATGAGATTTACCCGTGTGTAGAGGCGCATCTTCAGGTTTGTCTTCCTACAAAGTT 644	
Db	182 tagcctcaactgcttagtgtcttggtagacaatttgcctggcaccaagaataacgcaga 241	
QY	645 TACGCCCTAACTGGCTTAGTGTCTTGTGTAGACAAATTCCTCGGACCAAGATAACCGAGA 704	
Db	242 gagtgtgagaggataaaacggcaacacacagctcacagaacagactcttcagctgcgtgaa 301	
QY	705 GAGTGTAGAGAGGATAAACGGCAACACAGCTCACAGACACAGACATTCACAGTGTCTGAA 764	
Db	302 gttatggaacatcaaaacaaagaccgaagatatagttcaagaagaatcatccaaagattga 361	
QY	765 GTTATGGAAACATCAAAACAAAGACCAAGATATAGTCAAGAAGATCATCCAAAGATATGA 824	

JOURNAL  
COMMENT

Contact: Wilson RK  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

This clone is available royalty-free through LUNL; contact the  
 email: [esewatson@wustl.edu](mailto:esewatson@wustl.edu)  
 The IMAGE Consortium ([info@image.lunl.gov](mailto:info@image.lunl.gov)) for further information.  
 Seq primer: -28W13 rev2 from Amersham  
 High quality sequence stop: 258

1:50  
/organism="Homo sapiens"

```
/OryzaRsm- homo sapiens  
/note=vector; plusscript SK; Site_1: EcorI; Site_2:  
XhoI; Cloned unidirectionally. Primer: Oligo dT. HeLa S3  
epitheloid carcinoma cells grown to semi-confluency  
without induction. Average insert size: 1.5 kb; Uni-2AP XR  
Vector. -5' adaptor sequence: 5' GAATTCGCACGAG 3' -  
adaptor sequence: 5' CTCGACTGGTTCTTTTTTCTTGT 3'
```

```

/cclone="550527"
/cclone_lib="Stratagene HeLa cell s3 937216"
/sex="female"
/lab_host="HeLa S3 cell line"
/lab_host="SOLR (kanamycin resistant)"

```

ORIGIN	BASE COUNT	mRNA
1	1	U
2	1	U
3	1	U
4	1	U
5	1	U
6	1	U
7	1	U
8	1	U
9	1	U
10	1	U
11	1	U
12	1	U
13	1	U
14	1	U
15	1	U
16	1	U
17	1	U
18	1	U
19	1	U
20	1	U
21	1	U
22	1	U
23	1	U
24	1	U
25	1	U
26	1	U
27	1	U
28	1	U
29	1	U
30	1	U
31	1	U
32	1	U
33	1	U
34	1	U
35	1	U
36	1	U
37	1	U
38	1	U
39	1	U
40	1	U
41	1	U
42	1	U
43	1	U
44	1	U
45	1	U
46	1	U
47	1	U
48	1	U
49	1	U
50	1	U
51	1	U
52	1	U
53	1	U
54	1	U
55	1	U
56	1	U
57	1	U
58	1	U
59	1	U
60	1	U
61	1	U
62	1	U
63	1	U
64	1	U
65	1	U
66	1	U
67	1	U
68	1	U
69	1	U
70	1	U
71	1	U
72	1	U
73	1	U
74	1	U
75	1	U
76	1	U
77	1	U
78	1	U
79	1	U
80	1	U
81	1	U
82	1	U
83	1	U
84	1	U
85	1	U
86	1	U
87	1	U
88	1	U
89	1	U
90	1	U
91	1	U
92	1	U
93	1	U
94	1	U
95	1	U
96	1	U
97	1	U
98	1	U
99	1	U
100	1	U

BASE COUNT	ORIGIN	109 a	43 c	51 g	95 t	2 others
mRNA	<1..>300					

### h

2.0%;	Score 22;	DB 183;	Length 300;
Similarity 71.2%;	Pred. No. 1.45e-04;		
37: Conservative	0: Mismatches 15;	Indels 0;	Gaps 0;

Db            215   gtaacttcaattgaaccacgaaacctgtataatagcgtgtacagtgtaaagaagtgt   266

Ov            81   GTACCTTCATTTATGACCGAAGAAACCTCTCATCAGCTGTGTGTGTGCACAATGT   132

RESULT	5	RICC0437A - 330 bp mRNA EST	18-MAR-1995
LOCUS			
DEFINITION		Rice cDNA, partial sequence (C0437A).	

ACCESSION  
NID  
KEYWORDS  
SOURCE  
ORGANISM

D15309  
EST expressed sequence tag)  
9286502  
*Oryza sativa* (strain Nipponbare, ) callus cDNA to mRNA.  
*Oryza sativa*  
Eukaryotes; mitochondrial eukaryotes; Viridiplantae;  
Charophyta/Embryophyta group; Embryophyta; Magnoliophyta;  
Liliopsida; Poales; Poaceae; *Oryza*.

REFERENCE	AUTHORS	TITLE	JOURNAL
1	...	...	...
2	...	...	...
3	...	...	...
4	...	...	...
5	...	...	...
6	...	...	...
7	...	...	...
8	...	...	...
9	...	...	...
10	...	...	...
11	...	...	...
12	...	...	...
13	...	...	...
14	...	...	...
15	...	...	...
16	...	...	...
17	...	...	...
18	...	...	...
19	...	...	...
20	...	...	...
21	...	...	...
22	...	...	...
23	...	...	...
24	...	...	...
25	...	...	...
26	...	...	...
27	...	...	...
28	...	...	...
29	...	...	...
30	...	...	...
31	...	...	...
32	...	...	...
33	...	...	...
34	...	...	...
35	...	...	...
36	...	...	...
37	...	...	...
38	...	...	...
39	...	...	...
40	...	...	...
41	...	...	...
42	...	...	...
43	...	...	...
44	...	...	...
45	...	...	...
46	...	...	...
47	...	...	...
48	...	...	...
49	...	...	...
50	...	...	...
51	...	...	...
52	...	...	...
53	...	...	...
54	...	...	...
55	...	...	...
56	...	...	...
57	...	...	...
58	...	...	...
59	...	...	...
60	...	...	...
61	...	...	...
62	...	...	...
63	...	...	...
64	...	...	...
65	...	...	...
66	...	...	...
67	...	...	...
68	...	...	...
69	...	...	...
70	...	...	...
71	...	...	...
72	...	...	...
73	...	...	...
74	...	...	...
75	...	...	...
76	...	...	...
77	...	...	...
78	...	...	...
79	...	...	...
80	...	...	...
81	...	...	...
82	...	...	...
83	...	...	...
84	...	...	...
85	...	...	...
86	...	...	...
87	...	...	...
88	...	...	...
89	...	...	...
90	...	...	...
91	...	...	...
92	...	...	...
93	...	...	...
94	...	...	...
95	...	...	...
96	...	...	...
97	...	...	...
98	...	...	...
99	...	...	...
100	...	...	...

1 (bases 1 to 330)  
Sasaki, T. and Minobe, Y.  
Rice cDNA from callus  
Unpublished (1993)  
2 (sites)  
Sasaki, T., Song, J., Koga-Ban, Y., Matsui, E., Fang, F., Higo, H.,  
Nagasaki, H., Hori, M., Miya, M., Murayama-Kayano, E. et al.  
Toward cataloging all rice genes: large-scale sequencing of  
randomly chosen rice cDNAs from a callus cDNA library  
Plant J. 6 (4), 615-624 (1994)

MEDLINE	COMMENT
---------	---------

95078950  
Submitted (14-APR-1993) to DDBJ by:

Yuzo Minobe  
National Institute of Agrobiological Resources  
Rice Genome Research Program  
2-1-2 Kannondai, Tsukuba

Ibaraki,  
Japan 305  
Phone: 0298-38-7441  
Fax: 0298-38-7468  
PROJECT = 'RGP'

## FEATURES

Location/Qualifiers



Fax: 314 286 1810  
Email: est@watson.wustl.edu  
High quality sequence stops: 212  
Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

FEATURES  
Source  
1..410  
Location/Qualifiers  
/organism="Homo sapiens"  
/clone="85063"  
137 a 57 c 83 g 124 t 9 others

Query Match 2.0%; Score 22; DB 6; Length 410;  
Best Local Similarity 80.6%; Pred. No. 1.45e-04;  
Matches 29; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 163 aagatgatgaccagaagaatgaaattgtgaaac 198  
||||| ||| ||||||| ||| ||| |||||||  
QY 802 AAGAAGATCATCCAGATATTGACCTCTGTGANAAC 837

RESULT 9  
LOCUS  
DEFINITION  
SP:TCPE\_MOUSE P80314 T-COMPLEX PROTEIN 1, BETA SUBUNIT ;  
N21157 453 bp mRNA EST 19-DEC-1995  
Yx47D01.s1 Homo sapiens cDNA clone 264865 3', similar to  
ACCESSION N21157  
NID q1126327  
KEYWORDS  
SOURCE  
human clone=264865 primer=m13 -40 forward library=Soares melanocyte  
2NDHM vector=PT73D (Pharmacia) with a modified polylinker  
host=DRI10B (ampicillin resistant) Rsitel=Not I Rsite2=Eco RI Male.  
l1 strand cDNA was primed with a Not I - oligo(dT) primer  
[5'-TGTTACCACTCAAGTGGAGCGCGAGTGTGTGTGTGTGTGTGT-3'],  
double-stranded cDNA was size selected, ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I and Eco  
RI sites of a modified pT73 vector (Pharmacia). Library  
constructed by Bento Soares and M.Fatima Bonaldo. RNA from normal  
foreskin melanocytes (FS374) was kindly provided by Dr. Anthony P.  
Albino.

ORGANISM  
Homo sapiens  
Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;  
Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;  
Sarcopterygii; Chnoata; Tetrapoda; Amniota; Mammalia; Theria;  
Eutheria; Archonta; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 453)  
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,  
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,  
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,  
Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and  
Wilson,R.

REFERENCE  
AUTHORS  
The WashU-Merck EST Project  
Unpublished (1995)  
Contact: Wilton RK  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
High quality sequence stops: 321  
Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Possible reversed clone: similarity on wrong strand.  
Location/Qualifiers  
1..453  
/organism="Homo sapiens"  
/clone="264865"  
137 a 99 c 90 g 127 t

FEATURES  
Source  
1..453  
Location/Qualifiers  
/organism="Homo sapiens"  
/clone="264865"  
137 a 99 c 90 g 127 t

Query Match 2.0%; Score 22; DB 6; Length 410;  
Best Local Similarity 80.6%; Pred. No. 1.45e-04;  
Matches 29; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 163 aagatgatgaccagaagaatgaaattgtgaaac 198  
||||| ||| ||||||| ||| ||| |||||||  
QY 802 AAGAAGATCATCCAGATATTGACCTCTGTGANAAC 837

RESULT 9  
LOCUS  
DEFINITION  
SP:TCPE\_MOUSE P80314 T-COMPLEX PROTEIN 1, BETA SUBUNIT ;  
N21157 453 bp mRNA EST 19-DEC-1995  
Yx47D01.s1 Homo sapiens cDNA clone 264865 3', similar to  
ACCESSION N21157  
NID q1126327  
KEYWORDS  
SOURCE  
human clone=264865 primer=m13 -40 forward library=Soares melanocyte  
2NDHM vector=PT73D (Pharmacia) with a modified polylinker  
host=DRI10B (ampicillin resistant) Rsitel=Not I Rsite2=Eco RI Male.  
l1 strand cDNA was primed with a Not I - oligo(dT) primer  
[5'-TGTTACCACTCAAGTGGAGCGCGAGTGTGTGTGTGTGTGTGT-3'],  
double-stranded cDNA was size selected, ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I and Eco  
RI sites of a modified pT73 vector (Pharmacia). Library  
constructed by Bento Soares and M.Fatima Bonaldo. RNA from normal  
foreskin melanocytes (FS374) was kindly provided by Dr. Anthony P.  
Albino.

ORGANISM  
Homo sapiens  
Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;  
Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;  
Sarcopterygii; Chnoata; Tetrapoda; Amniota; Mammalia; Theria;  
Eutheria; Archonta; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 453)  
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,  
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,  
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,  
Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and  
Wilson,R.

REFERENCE  
AUTHORS  
The WashU-Merck EST Project  
Unpublished (1995)  
Contact: Wilton RK  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
High quality sequence stops: 321  
Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Possible reversed clone: similarity on wrong strand.  
Location/Qualifiers  
1..453  
/organism="Homo sapiens"  
/clone="264865"  
137 a 99 c 90 g 127 t

FEATURES  
Source  
1..453  
Location/Qualifiers  
/organism="Homo sapiens"  
/clone="264865"  
137 a 99 c 90 g 127 t

Query Match 2.0%; Score 22; DB 6; Length 410;  
Best Local Similarity 80.6%; Pred. No. 1.45e-04;  
Matches 29; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 163 aagatgatgaccagaagaatgaaattgtgaaac 198  
||||| ||| ||||||| ||| ||| |||||||  
QY 802 AAGAAGATCATCCAGATATTGACCTCTGTGANAAC 837

RESULT 9  
LOCUS  
DEFINITION  
SP:TCPE\_MOUSE P80314 T-COMPLEX PROTEIN 1, BETA SUBUNIT ;  
N21157 453 bp mRNA EST 19-DEC-1995  
Yx47D01.s1 Homo sapiens cDNA clone 264865 3', similar to  
ACCESSION N21157  
NID q1126327  
KEYWORDS  
SOURCE  
human clone=264865 primer=m13 -40 forward library=Soares melanocyte  
2NDHM vector=PT73D (Pharmacia) with a modified polylinker  
host=DRI10B (ampicillin resistant) Rsitel=Not I Rsite2=Eco RI Male.  
l1 strand cDNA was primed with a Not I - oligo(dT) primer  
[5'-TGTTACCACTCAAGTGGAGCGCGAGTGTGTGTGTGTGTGTGT-3'],  
double-stranded cDNA was size selected, ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I and Eco  
RI sites of a modified pT73 vector (Pharmacia). Library  
constructed by Bento Soares and M.Fatima Bonaldo. RNA from normal  
foreskin melanocytes (FS374) was kindly provided by Dr. Anthony P.  
Albino.

ORGANISM  
Homo sapiens  
Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;  
Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;  
Sarcopterygii; Chnoata; Tetrapoda; Amniota; Mammalia; Theria;  
Eutheria; Archonta; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 453)  
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,  
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,  
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,  
Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and  
Wilson,R.

REFERENCE  
AUTHORS  
The WashU-Merck EST Project  
Unpublished (1995)  
Contact: Wilton RK  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
High quality sequence stops: 321  
Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Possible reversed clone: similarity on wrong strand.  
Location/Qualifiers  
1..453  
/organism="Homo sapiens"  
/clone="264865"  
137 a 99 c 90 g 127 t

FEATURES  
Source  
1..453  
Location/Qualifiers  
/organism="Homo sapiens"  
/clone="264865"  
137 a 99 c 90 g 127 t

Query Match 2.0%; Score 22; DB 6; Length 410;  
Best Local Similarity 80.6%; Pred. No. 1.45e-04;  
Matches 29; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 163 aagatgatgaccagaagaatgaaattgtgaaac 198  
||||| ||| ||||||| ||| ||| |||||||  
QY 802 AAGAAGATCATCCAGATATTGACCTCTGTGANAAC 837

RESULT 9  
LOCUS  
DEFINITION  
SP:TCPE\_MOUSE P80314 T-COMPLEX PROTEIN 1, BETA SUBUNIT ;  
N21157 453 bp mRNA EST 19-DEC-1995  
Yx47D01.s1 Homo sapiens cDNA clone 264865 3', similar to  
ACCESSION N21157  
NID q1126327  
KEYWORDS  
SOURCE  
human clone=264865 primer=m13 -40 forward library=Soares melanocyte  
2NDHM vector=PT73D (Pharmacia) with a modified polylinker  
host=DRI10B (ampicillin resistant) Rsitel=Not I Rsite2=Eco RI Male.  
l1 strand cDNA was primed with a Not I - oligo(dT) primer  
[5'-TGTTACCACTCAAGTGGAGCGCGAGTGTGTGTGTGTGTGTGT-3'],  
double-stranded cDNA was size selected, ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I and Eco  
RI sites of a modified pT73 vector (Pharmacia). Library  
constructed by Bento Soares and M.Fatima Bonaldo. RNA from normal  
foreskin melanocytes (FS374) was kindly provided by Dr. Anthony P.  
Albino.

ORGANISM  
Homo sapiens  
Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;  
Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;  
Sarcopterygii; Chnoata; Tetrapoda; Amniota; Mammalia; Theria;  
Eutheria; Archonta; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 453)  
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,  
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,  
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,  
Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and  
Wilson,R.

REFERENCE  
AUTHORS  
The WashU-Merck EST Project  
Unpublished (1995)  
Contact



DEFINITION  
H. sapiens partial cDNA sequence; clone c-0be06.  
ACCESSION  
238433  
NID  
9560441  
KEYWORDS  
partial cDNA sequence; transcribed sequence fragment.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryotae; Mitochondrial eukaryotes; Metazoa; Chordata;  
Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 253)  
Genexpress.  
AUTHORS  
Direct Submission  
TITLE  
Submitted (24-OCT-1994) Genethon, B.P. 60, 91002 Evry Cedex France  
JOURNAL  
and Genetique Moleculaire et Biologie du developpement, CNRS UPR420  
B.P. 8, 94801 Villejuif Cedex France.E-mail: genexpress@genethon.fr  
REFERENCE  
2 (bases 1 to 253)  
Genexpress.  
AUTHORS  
The Genexpress cDNA program  
TITLE  
Unpublished  
JOURNAL  
3 (bases 1 to 253)  
Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C.,  
AUTHORS  
Devignes,M.D., Duprat,S., Houligatte,R., Juneau,M.N., Lamy,B.,  
Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,  
Sebastiani-Kabakchis,C. and Tessier,A.  
TITLE  
IMAGE: molecular integration of the analysis of the human genome  
JOURNAL  
and its expression  
MEDLINE  
C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)  
COMMENT  
95277534  
Clone library from B.Souares, Psychiatry Dept. Columbia University  
USA;

Cloning\_method: total mRNA was oligo-(dT) primed and directionally  
cloned 5' -> 3' into the HindIII -> NotI sites of the lafmid BA  
vector;  
Sequencing\_method: single read, full automatic;  
Primer: (-21)M13 universal;  
cDNA sequence complementary to mRNA (3'end)  
Stretch\_removed: 31 T removed at sequence 5'end  
Normalization\_method: Bento Soares, P.N.A.S in press;  
Genexpress\_library\_id: C;  
Genexpress\_sequence\_id: alc-0be06;

No significant homology found with :  
genbank release 81 swissprot release 28.

#### FEATURES

Source  
1..253  
Location/Qualifiers  
/organism="Homo sapiens"  
/isolate="muscular atrophy patient"  
/dev\_stage="3 months old"  
/tissue\_type="total brain"  
/clone\_lib="normalized infant brain cDNA"  
/sex="Female"  
BASE COUNT 76 a 57 c 49 g 70 t 1 others  
ORIGIN

Query Match 1.9%; Score 21; DB 129; Length 253;  
Best Local Similarity 78.4%; Pred. No. 2.90e-03;  
Matches 29; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Db 123 acacagctcacatgtacacacataaaactgtctcaag 159  
|||||  
QY 729 ACACAGCTCACAGACAGACTTTCACGCTGCTGAAG 765

RESULT 15  
LOCUS T89287 265 bp mRNA EST 20-MAR-1995  
DEFINITION yd37f05.s1 Homo sapiens cDNA clone 110433 3'.

ACCESSION T89287  
NID 9717800  
KEYWORDS EST.  
SOURCE human clone-110433 library-Soares fetal liver spleen 1NFLS  
vector-pT7T3D (Pharmacia) with a modified polylinker host=DH10B  
(ampicillin resistant) primer=-21ml3 Rsite1-Pac I Rsite2-Eco RI  
Liver and spleen from a 20 week-post conception male fetus. 1st

strand cDNA was primed with a Pac I - oligo(dT) primer [5',  
AACTGGAAGATTAATAAGATCTTTTTTTTTTTTTTTT 3'], double-stranded  
cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac  
I and cloned into the Pac I and Eco RI sites of the modified pT7T3  
vector. Library went through one round of normalization. Library  
constructed by Bento Soares and M.Fatima Bonaldo.  
Homo sapiens  
Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 265)  
AUTHORS  
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,  
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,  
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,  
Trevasakis,E., Waterston,R., Williamson,A., Wohldmann,P. and  
Wilson,R.  
TITLE  
The WashU-Merck EST Project  
JOURNAL  
Unpublished (1995)  
COMMENT

Contact: Wilson RK  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

High quality sequence stops: 241

Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

#### FEATURES

source  
1..265  
Location/Qualifiers  
/organism="Homo sapiens"  
/clone="110433"

BASE COUNT 82 a 51 c 68 g 47 t 7 others  
ORIGIN

Query Match 1.9%; Score 21; DB 9; Length 265;  
Best Local Similarity 66.7%; Pred. No. 2.90e-03;  
Matches 36; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Db 77 aaaaagacagagagacnaaagaagacatcagctctctccactcttccca 130  
|||||  
QY 958 AAAACTCTCACTCAGAGTCTAAAGAAGACCATCAGGTCTCTTCACAGCTTCA 1011

Search completed: Tue Dec 2 16:51:11 1997  
Job time : 621 secs.





\*\*\*\*\*

MAISELH  
(TM)

\*\*\*\*\*

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelligentGenetics, Inc.

MPsrch\_nnn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 2 16:36:04 1997; MasPar time 121.52 Seconds  
935.265 Million cell updates/sec

file: >US-08-915-004-10

Description: (1-1089) from US08915004.seq

Perfect Score: 1089

N.A. Sequence:

Comp: 1 ATGAACAAGTTGCTGTGCTG.....TAAAAATAAGCTGTATAA 1089

TACTTGTTCACGACACGAC.....ATTITTTCGACGAATTT

Scoring table: TABLE default

Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 142080 seqs, 52183452 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: n-geneseq28

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29

Statistics: Mean 9.038; Variance 4.941; scale 1.829

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	ID	Description	Pred. No.
1	871	80.0	1173	28	Human tumour necrosis	0.00e+00
2	869	79.8	966	27	Mutated OCIF, OCIF-CS	0.00e+00
3	869	79.8	1056	27	Mutated OCIF, OCIF-CC	0.00e+00
4	869	79.8	1182	27	Mutated OCIF, OCIF-CB	0.00e+00
5	869	79.8	1200	27	Mutated OCIF, OCIF-CL	0.00e+00
6	869	79.8	1206	27	Mutated OCIF, OCIF-C2	0.00e+00
7	869	79.8	1206	28	Osteoclastogenesis in	0.00e+00
8	867	79.6	1206	27	Mutated OCIF, OCIF-C2	0.00e+00
9	865	79.4	1206	28	Mutated OCIF, OCIF-C2	0.00e+00
10	865	79.4	1206	27	Mutated OCIF, OCIF-C2	0.00e+00
11	865	79.4	1206	28	Mutated OCIF, OCIF-C1	0.00e+00
12	816	74.9	984	27	Mutated OCIF, OCIF-DD	0.00e+00
13	814	74.7	819	27	Mutated OCIF, OCIF-CD	0.00e+00
14	682	62.6	1083	27	Mutated OCIF, OCIF-DC	0.00e+00
15	598	54.9	981	27	Mutated OCIF, OCIF-DD	0.00e+00

16	589	54.1	594	27	T33175	Mutated OCIF, OCIF-CD	0.00e+00
17	558	51.2	564	27	T33180	Mutated OCIF, OCIF-CB	0.00e+00
18	557	51.1	1080	27	T33167	Mutated OCIF, OCIF-DC	0.00e+00
19	484	44.4	1080	27	T33169	Mutated OCIF, OCIF-DC	0.00e+00
20	442	40.6	1080	27	T33168	Mutated OCIF, OCIF-DC	0.00e+00
21	428	39.3	432	27	T33176	Mutated OCIF, OCIF-CC	0.00e+00
22	402	36.9	438	28	T36689	Osteoclastogenesis in	7.84e-283
23	400	36.7	465	28	T36688	Osteoclastogenesis in	2.81e-281
24	373	34.3	10190	27	T33183	Fragment of human OCI	2.67e-260
25	318	29.2	321	27	T33177	Mutated OCIF, OCIF-CC	1.10e-217
26	244	22.4	255	27	T33181	Mutated OCIF, OCIF-CP	1.10e-160
27	82	7.5	1047	2	Q10572	Human Natriuretic Pep	2.55e-39
28	64	5.9	1047	2	Q10572	Human Natriuretic Pep	1.12e-26
29	45	4.1	91	9	Q51746	Oligonucleotide probe	6.20e-14
30	42	3.9	204	1	N81164	Base substituted E.co	5.22e-12
31	39	3.6	91	9	Q51746	Oligonucleotide probe	4.06e-10
32	37	3.4	204	1	N81164	Base substituted E.co	7.06e-09
33	36	3.3	114	12	Q70470	Generic DNA sequence	2.90e-08
34	36	3.3	114	12	Q70469	Generic DNA sequence	2.90e-08
35	36	3.3	114	12	Q70465	Generic DNA sequence	2.90e-08
36	35	3.2	114	12	Q70467	Generic DNA sequence	1.17e-07
37	34	3.1	114	12	Q70467	Generic DNA sequence	4.71e-07
38	33	3.0	114	12	Q70468	Generic DNA sequence	1.86e-06
39	33	3.0	114	12	Q70466	Generic DNA sequence	1.86e-06
40	33	3.0	114	12	Q70468	Generic DNA sequence	1.86e-06
41	32	2.9	114	12	Q70473	Generic DNA sequence	7.26e-06
42	32	2.9	114	12	Q70469	Generic DNA sequence	7.26e-06
43	31	2.8	114	12	Q70465	Generic DNA sequence	7.26e-06
44	31	2.8	114	12	Q70472	Generic DNA sequence	2.79e-05
45	30	2.8	114	12	Q70471	Generic DNA sequence	1.06e-04

## ALIGNMENTS

RESULT 1  
ID T35475 standard; cDNA; 1173 BP.

AC T35475;  
DT 05-MAY-1997 (first entry)  
DE Human tumour necrosis factor receptor.  
KW Tumour necrosis factor; TNF; receptor; TNF-beta; ligand; tumour;  
KW differentiation; immune response; autoimmune disease; inflammation;  
KW septic shock; graft-versus-host; apoptosis; ss.  
FH Homo sapiens.  
FT sig\_peptide 1..63 Location/Qualifiers  
FT /tag= a  
FT CDS 1..1173  
FT /\*tag= b  
PN WO9628546-A1.  
PD 19-SEP-1996.  
PF 15-MAR-1995; U03216.  
PR 15-MAR-1995; WO-U03216.  
PR 29-MAR-1995; ZA-002587.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Fleischmann RD, Greene JM;  
DR WPI: 96-433821/43.  
DR P-PDB: R9Q357.  
PT New human tumour necrosis factor receptor - used to develop prods.  
PT for treating e.g. tumours, infection, auto-immune disease, graft  
PT rejection, cytotoxicity or inflammation  
PS Claim 1; Fig 1; 59pp; English.  
CC The receptor binds to TNF, and in particular, TNF-beta.  
CC The receptor may be used for screening for antagonists and agonists  
CC of the receptor and for ligands for the receptor. Such agonists may  
CC be used to inhibit the growth of tumours, to stimulate cellular  
CC differentiation, to mediate the immune response and anti-viral  
CC response, to regulate growth and provide resistance to certain  
CC infections. The antagonists may be used therapeutically, to treat  
CC autoimmune diseases, inflammation, septic shock, to inhibit graft-  
CC versus-host reactions, and to prevent apoptosis.  
SQ Sequence 1173 BP; 374 A; 277 C; 265 G; 257 T;

Query Match 80.0%; Score 871; DB 28; Length 1173;





PR 21-JUL-1995; JP-207508.  
 PA (SNOW) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR P-PSDB: R99948.  
 PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 78: Page 148; 183pp; Japanese.  
 CC This sequence encodes a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC sequence encodes OCIF-CBst in which Gln371 is substituted with Leu  
 CC and amino acids 373-380 of the mature protein have been deleted. These  
 CC amino acid changes have been caused by the introduction of a restriction  
 CC site. The OCIF of the invention has a molecular weight by SDS-PAGE of  
 CC 60 kD under reducing conditions and 120 kD under non-reducing  
 CC conditions. The protein is adsorbed onto cation-exchangers or heparin  
 CC and its activity is lowered after 10 mins at 70 deg.C or 30 mins at 56  
 CC deg.C, and is lost after 10 mins at 90 deg.C. OCIF is useful in the  
 CC control of bone resorption and therefore in the treatment and  
 CC prevention of disorders of bone resorption, e.g. osteoporosis.  
 CC Sequence 1182 BP; 376 A; 280 C; 266 G; 260 T;  
 SQ

Query Match 79.8%; Score 869; DB 27; Length 1182;  
 Best Local Similarity 99.9%; Pred. No. 0.00e+00;  
 Matches 870; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 atgaacaactgctgctgcgcgcctcgtgtttctgacatctccattaaagtggaccacc 60  
 Qy 1 ATGACAAGTTGCTGTGCTGCGCGCTCGTGTTCCTGGACATCTCCATTAAAGTGACCA 60  
 Db 61 caggaaacgtttctccaaagtacattcattatgacgaagaacacctctcatcagctgttg 120  
 Qy 61 CAGGAACGTTTCTCCAAAGTACCTTCATTATGACGAAGAAGAACTCTCATCAGCTGTG 120  
 Db 121 tggacaaatctctcctggctacactactacacagacagctggacacagctgacagctgt 180  
 Qy 121 TGTGACAAATGCTCTCTGTTACTTACTTAAACAACTGACGACAAAGTGGAGAAC 180  
 Db 181 gttgcccccttccctggaccactactacacagacagctggacacagctgacagctgt 240  
 Qy 181 GTGTGCGCCCTTCCCTGGACCTACTACACAGACGCTGGACATCTCCATTAAAGTGACCA 240  
 Db 241 ctatactgagccctcctgctgcaagagctgcagctacgtcaagcaggagtgcaatgcacc 300  
 Qy 241 CTATACTGACCCCGCTGTGCAAGGAGCTGCAAGTACGTCAAGCAGGAGTGCATTCGCA 300  
 Db 301 cacaaccgctgtgcgaatgcaaggaaggcgctaccttgagatagagcttctgcttgaaa 360  
 Qy 301 CACAACCCGTGTGGGAATGCAAGGAAGGCGCTACCTTGAGATAGAGTCTGCTTGAAA 360  
 Db 361 cataggagctgccccctcctggatttggagtggtgcaagctggaaccccgagagcgaaataca 420  
 Qy 361 CATAGGAGCTGCCCTCCTCGGATTTGGAGTGTGCAAGCTGGAAACCCGACGAGCAATACA 420  
 Db 421 gttgcaaaagatgctcagatgggtttcttcaaatgagacgctcatctaagcaccctgt 480  
 Qy 421 GTTTGCAAAAGATGCTCAGATGGGTTTCTTCAAAATGAGAGCTCATCTAAAGCACCCCTGT 480  
 Db 481 agaaacacacaaattgacgtgtcttggctcctcgttaactcagaaggaatgcaaca 540  
 Qy 481 AGAAACACACAAATTGACGTGTCTTTGGTCTCCTGCTACTCAGAAAGGAATGCAACA 540  
 Db 541 cagcaacacatatttccggaacagtgaaatcaactcaaaaatggaatagatgttacc 600  
 Qy 541 CACGACACATATGTTCCGGAACAGTGAATCAACTCAAAATGTGGAATAGATGTACC 600  
 Db 601 ctgtgtgaggaggtattctcagttgctgttctcctcaaaagttagcctaactggctt 660  
 Qy 601 CTGTGTGAGGAGGATTCTTCTCAGGTTGCTGTCTTCTCAAAAGTTTACGCCCTAACTGCGCTT 660  
 Db 661 agtgtctgttagacaatttgcctggcaccacaaagtaaacgcagagagtgtagagagata 720  
 Qy 661 GTGTGCGCCCTTGGCCCTGACCACTACTACTACACAGACGCTGGACACACGAGTGGAGTGT 720

Qy 661 AGTGTCTTGGTAGACAATTTGCCCTGGCACCACAAAGTAACACGACAGAGTGTAGAGAGATA 720  
 Db 721 aaacggcaacacagctcacagaacacagacttccagctgtgagttatgaaacatcaa 780  
 Qy 721 AAACGGCAACACAGCTCACAGAACACAGACTTCCAGCTGTGAGTTATGGAACATCAA 780  
 Db 781 aacaaagacacaaagtatactagtcacaaagatcatcacaagattatgacctctgtgaaacagc 840  
 Qy 781 AACAAAGACCAAGATATAGTCAAGAAGATCATCCAGATATTGACCTCTGTGAAACACAGC 840  
 Db 841 gtgacggcgacattggacatgctaaactca 871  
 Qy 841 GTGCAGCGGCACATTGGACATGCTAACTCA 871

RESULT 5  
 ID T33172 standard; DNA; 1200 BP.  
 AC T33172, 1997 (first entry)  
 DE Mutated OCIF, OCIF-CL, coding sequence.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 KW osteoporosis; ss.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT sig\_peptide 1..63  
 FT /\*tag= a  
 FT mat\_peptide 64..1197  
 FT /\*tag= b  
 FT /\*product= OCIF-CL  
 PN W09626217-A1.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; J00374.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR P-PSDB: R99948.  
 PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 60: Page 143-144; 183pp; Japanese.  
 CC This sequence encodes a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC sequence encodes OCIF-CL in which amino acids 379-380 of the mature  
 CC protein have been deleted. The OCIF of the invention has a molecular  
 CC weight by SDS-PAGE of 60 kD under reducing conditions and 120 kD under  
 CC non-reducing conditions. The protein is adsorbed onto cation-exchangers  
 CC or heparin and its activity is lowered after 10 mins at 70 deg.C or 30  
 CC mins at 56 deg.C, and is lost after 10 mins at 90 deg.C. OCIF is useful  
 CC in the control of bone resorption and therefore in the treatment and  
 CC prevention of disorders of bone resorption, e.g. osteoporosis.  
 CC Sequence 1200 BP; 387 A; 283 C; 268 G; 262 T;  
 SQ

Query Match 79.8%; Score 869; DB 27; Length 1200;  
 Best Local Similarity 99.9%; Pred. No. 0.00e+00;  
 Matches 870; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 atgaacaactgctgctgcgcgcctcgtgtttctgacatctccattaaagtggaccacc 60  
 Qy 1 ATGACAAGTTGCTGTGCTGCGCGCTCGTGTTCCTGGACATCTCCATTAAAGTGACCA 60  
 Db 61 caggaaacgtttctccaaagtacattcattatgacgaagaacacctctcatcagctgttg 120  
 Qy 61 CAGGAACGTTTCTCCAAAGTACCTTCATTATGACGAAGAAGAACTCTCATCAGCTGTG 120  
 Db 121 tggacaaatgctcctcctggctacactactacacacacactacacagaaagtgaagacc 180  
 Qy 121 TGTGACAAATGCTCTCTGTTACTTACTTAAACAACTGACGACAAAGTGGAGAAC 180  
 Db 181 gttgcccccttccctggaccactactacacagacagctggacacagctgacagctgt 240  
 Qy 181 GTGTGCGCCCTTGGCCCTGACCACTACTACTACACAGACGCTGGACATCTCCATTAAAGTGACCA 240



Db 841 gtgacgagccacattgagacatgctaactca 871  
 QY 841 GTGCAGCGCACATTGGACATGCTAACTCA 871

RESULT 7  
 ID T33685 standard; DNA; 1206 BP.  
 AC T33685;  
 DT 22-APR-1997 (first entry)  
 DE Osteoclastogenesis inhibitory factor coding sequence.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 KW osteoporosis; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT sig\_peptide 1..63  
 FT /\*tag= a  
 FT mat\_peptide 64..1203  
 FT /\*tag= b  
 /label= Claim 6  
 /WO9626217-A1.  
 PD 29-AUG-1996.  
 PF 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR P-SDS: 96-402320/40.  
 DR P-PSDB: R99924-25.  
 PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 8; Page 66-67; 183pp; Japanese.  
 CC This sequence encodes the full length osteoclastogenesis inhibitory  
 CC factor (OCIF) of the invention. The OCIF has a molecular weight by  
 CC SDS-PAGE of 60 kD under reducing conditions and 120 kD under non-  
 CC reducing conditions. The protein is adsorbed onto cation-exchangers  
 CC or heparin and its activity is lowered after 10 mins at 70 deg.C or  
 CC 30 mins at 56 deg.C, and is lost after 10 mins at 90 deg.C. OCIF is  
 CC useful in the control of bone resorption and therefore in the  
 CC treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 SQ Sequence 1206 BP; 388 A; 284 C; 269 G; 265 T;

Query Match 79.8%; Score 869; DB 28; Length 1206;  
 Best Local Similarity 99.9%; Pred. No. 0.00e+00;  
 Matches 870; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 atgaacactgctgtgctgcgcctgctgttcttgacatctccattgaagtgaccac 60  
 |||||  
 1 ATGAACAAGTTGCTGTGCTGCGCGCTGTGTCTTGACATCTCCATTAAGTGGACACC 60  
 |||||

61 caggaaacgtttctctcctcaaaagtacattcattatgacgaagaacacctctcagctgtg 120  
 |||||  
 61 CAGGAACGTTTCTCTCCAAAGTACCTTCAITATACGAAGAAGAACTCTCATCAGCTGTG 120  
 |||||

121 tgtacaaatctctctctgtgtacctactataaacaacactgtacagcaagtgtgaagacc 180  
 |||||  
 121 TGTACAAATCTCTCTGTGTACCTACTCTAAACAACACTGTACAGCAAAAGTGGAGACC 180  
 |||||

181 gtgtgcgcccccttcccttgaccactactacacagacgtggcacaccagtgcaggtgt 240  
 |||||  
 181 GTGTGCGCCCCCTTCCCTTGACCACTACTACAGACAGCTGGCACCACTGACGAGTGT 240  
 |||||

241 ctatactcagccccctgtgcaagagctgcagtcacgtcgaagcaggtgcgaatcgcaacc 300  
 |||||  
 241 CTATACTCAGCCCCGTGTGCAAGAGCTGCAGTACGTCACGACGAGTGCATCGCACC 300  
 |||||

301 cacaaccgctgtgcgaatgcgaaggaagggcgctaccttagatagatagattctcttgaaa 360  
 |||||  
 301 CACAACC CGTGTGCGAATGCGAAGGAAGGGCGCTACCTTGTAGATAGATTCTGCTGAAA 360  
 |||||

361 cataggagctccctctctgtgattggagtgtgcaagctggaaccccgacgcaaatata 420  
 |||||

QY 361 CATAGGAGCTGCCCTCTCTGGATTGGAGTGTGCAAGCTGGAAACCCAGAGCGAAATACA 420  
 Db 421 gtttgcacaaatgtccagatgggttctctcctcaaatgagacgtcatctaaagcaccctgt 480  
 |||||  
 QY 421 GTTTGCAAAAGATGTCCAGATGGGTCTTCTCAATGAGAGCTCATCTAAAGCACCTGT 480  
 |||||

Db 481 agaaaacacacaaattgacgtgtcttctgctctctgtaactcagaagaagaaatgcacaa 540  
 |||||  
 QY 481 AGAAAACACACAAATTGCAGTGTCTTTGGTCTCTGCTAACTCAGAAAGGAAATGCAACA 540  
 |||||

Db 541 cagcacaacatatgttcoggaacacagtgaaatcaactcaaaaatgtggaatagatttacc 600  
 |||||  
 QY 541 CACGACAACATATGTTCCGGAAACAGTGAATCAACTCAAAATGTGGAATAGATTGTACC 600  
 |||||

Db 601 ctgtgtgagggaggtattcttcagggtttgctgttctctcaaaagtttacgcttaactggcctt 660  
 |||||  
 QY 601 CTGTGTGAGGGAGGATCTTTCAGGTTTGCTGTTCTTACAAAGTTTACGCTTAATGGCTT 660  
 |||||

Db 661 agtctcttggtagacaatttgcctggcaccacaaagtaaacgcagagagtgtagagagata 720  
 |||||  
 QY 661 AGTGTCTTGGTAGACAATTTCCTGGCACCACAAAGTAAACGCAGAGAGTGTAGAGAGGATA 720  
 |||||

Db 721 aaacggcaacacagctcacaagaacagacttccagctgtggaagttatggaaacataca 780  
 |||||  
 QY 721 AAACGGCAACACAGCTCACAAAGACAGACTTTCAGCTGTCTGAAAGTTATGGAACATCAA 780  
 |||||

Db 781 aacaagaccacagatatagtcacgaagaatcatccaagatatgtacctctgtgaaacacagc 840  
 |||||  
 QY 781 AACAAAGACCAAGATATAGTCAAGAAGATCATCCAAGATATTGACCTCTGTGAAAACAGC 840  
 |||||

Db 841 gtgcgagcgacattggacatgctaactca 871  
 |||||  
 QY 841 GTGCAGCGCACATTGGACATGCTAACTCA 871  
 |||||

RESULT 8  
 ID T33165 standard; DNA; 1206 BP.  
 AC T33165;  
 DT 22-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-C23S, coding sequence.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 KW osteoporosis; ss.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT sig\_peptide 1..63  
 FT /\*tag= a  
 FT mat\_peptide 64..1203  
 FT /\*tag= b  
 FT /product= OCIF-C23S  
 PN WO9626217-A1.  
 PD 29-AUG-1996.  
 PF 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR P-PSDB: R99935.  
 PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 39; Page 136-137; 183pp; Japanese.  
 CC This sequence encodes a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC sequence encodes OCIF-C23S in which the 23rd Cys residue in the mature  
 CC OCIF protein is substituted by Ser. The OCIF of the invention has a  
 CC molecular weight by SDS-PAGE of 60 kD under reducing conditions and  
 CC 120 kD under non-reducing conditions. The protein is adsorbed onto  
 CC cation-exchangers or heparin and its activity is lowered after 10 mins  
 CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
 CC deg.C. OCIF is useful in the control of bone resorption and therefore  
 CC in the treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.



SQ Sequence 1206 BP; 389 A; 285 C; 268 G; 264 T;  
 Query Match 79.6%; Score 867; DB 27; Length 1206;  
 Best Local Similarity 99.8%; Pred. No. 0.00e+00;  
 Matches 869; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db	1	atgaacaacttgctgtgctgcgcgcgtgtgtttcttgacatctcattaaatgacacc	60
QY	1	ATGAACAAGTTCGTGCTCGCGGCGTCGTGTTCTGGACATCTCCATTAAATGGACACC	60
Db	61	csggaacgtttcttccaaagtactcttattgaagaagaacctctcatcagctgttg	120
QY	61	CAGGAACGTTTCCTCCAAAGTACCTTCATTATGAGGAAGAAACCTCTCATCAGCTGTG	120
Db	121	ttgcaaatgtctctctctgtgtacctaataaacaacactgtacagaanaatggaagacc	180
QY	121	TGTGCAAAATGCTCTCTGTGTACTACTCTAAAACAACACTGTACAGCAAAATGGGAAGACC	180
Db	181	gtgtgcgccttgccttgaccactactacacagacagctggcacaccagtgacagtgct	240
QY	181	GTGTGCGCCCTTGCCCTGACCACTACTACAGACAGAGCTGGCACACAGTGCAGAGTGT	240
Db	241	ctatactgcagcccgctgtgcaaggagctgcactcaactcaacacactgtcaacagcagagtgaatgcacc	300
QY	241	CTATACTGCAGCCCGCTGTGCAAGGAGCTGCAGTACTCAAGCAGGAGTGCATTCGCACC	300
Db	301	cacaaccgctgtgcgaatgcaagaaggcgcgtactcttgatagatagagttctgttgaaa	360
QY	301	CACACCGCGTGTGCGAATGCAAGAAGGGCGTACTCTTGAGTAGAGTCTTGCTTGAAA	360
Db	361	cataggagctgcctctctggatttggagtggtgcaagctggaaacccacagacgcaataca	420
QY	361	CATAGGAGCTGCCCTCCCTGGATTGTGGAGTGGTGTCAAGCTGGAACCCACAGAGCAATACA	420
Db	421	gtttgcaaaagatgtccagatgtgggttcttctcaaatgagacgtcatcctaaagcacctgt	480
QY	421	GTTTGCAAAAGATGTCCAGATGGGTTCTTCTCAAATGAGACGTCACTAAAGCACCCCTGT	480
Db	481	agaaacacacaaaatgcagtgcttttggctctctctgctaactcagaaaggaatgcaaca	540
QY	481	AGAAACACACAAATTCAGTGTCTTTGGTCTCTCTTAACCTCAGAAAGGAAATGCAACA	540
Db	541	cacgcaacacatgtcttcgggaaacagtgaaatcaactcaaaaatgtggaatagattacc	600
QY	541	CAGCAACAACATATGTTCCGGAACAGTGAATCAACTCAAAAATGTGGAATAGATGTTACC	600
Db	601	ctgtgtgaggagcattcttcaggtttgtcttctcacaagtttacgcctaacctggctt	660
QY	601	CTGTGTGAGGAGGCATTCTTCAGGTTTGCTGTTCTCTACAAGTTTACGCTTAACCTGGCTT	660
Db	661	agtgctttggtagacaatttgcctgpgcaccaaaagttaaacycagagagttagagaggata	720
QY	661	AGTGCTTTGTTAGACAATTTGGCTGGCACCAAGTAAACAGAGAGTGTAGAGAGGATA	720
Db	721	aaacggcaacacagctcacaagaacagactttccagctgctgaagttatggaaacatcaa	780
QY	721	AAACGGCAACACAGCTCACAGAACAGACTTTCACGCTGCTGAAGTTATGGAAACATCA	780
Db	781	aacaaagaccagatatagtcaagaagatcatccaagatatgacctctgtgaaacagc	840
QY	781	AACAAAGACCAAGATATAGTCAAGAAGATCATCCAAGATATTGACCTCTGTGAAACAGC	840
Db	841	gtgcagcggcaattggtgacatgctaactca	871
QY	841	GTGCAGCGGCACATTGGACATGTAACCTCA	871

RESULT 9  
 ID T33162 standard; DNA; 1206 BP.  
 AC T33162;  
 DT 22-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-C20S, coding sequence.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;

KW	osteoporosis; ss.
OS	Synthetic.
FH	Key
FT	Location/Qualifiers
FT	sig_peptide
FT	/tag= a
FT	mat_peptide
FT	/tag= b
FT	/product= OCIF-C20S
PN	W09626217-A1.
PD	28-AUG-1996.
FF	20-FEB-1996; J00374.
FF	20-FEB-1995; JP-054977.
PR	21-JUL-1995; JP-207508.
PA	(SNOW ) SNOW BRAND MILK PROD CO LTD.
PI	Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;
PI	Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;
DR	WPI: 96-402320/40.
DR	P-PSDB: R99932
DR	DNA encoding osteoclastogenesis inhibitory factor protein - useful
PT	for bone resorption control, esp. treatment of osteoporosis
PS	Claim 30; Page 133-134; 183pp; Japanese.
CC	This sequence encodes a mutated version of the full length
CC	osteoclastogenesis inhibitory factor (OCIF) of the invention. This
CC	sequence encodes OCIF-C20S in which the 20th Cys residue in the mature
CC	OCIF protein is substituted by Ser. The OCIF of the invention has a
CC	molecular weight by SDS-PAGE of 60 kD under reducing conditions and
CC	120 kD under non-reducing conditions. The protein is adsorbed onto
CC	cation-exchangers or heparin and its activity is lowered after 10 mins
CC	at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90
CC	deg.C. OCIF is useful in the control of bone resorption and therefore
CC	in the treatment and prevention of disorders of bone resorption, e.g.
CC	osteoporosis.
SC	Sequence 1206 BP; 389 A; 283 C; 270 G; 264 T;

Query Match 79.4%; Score 865; DB 28; Length 1206;  
 Best Local Similarity 99.7%; Pred. NO. 0.00e+00;  
 Matches 869; Conservative 3; Indels 0; Gaps 0

Ddb  
Qy

1 atgaacaacttgctgtgcgcgctcggttttctggacatctccattaagtggaccacg 60  
|||||  
1 ATGACACAGTTGCTGTGCTGC GCGTGGTTTTCTGGACATCTCCATTAA GTGGACCACC 60

Ddb  
Qy

61 caggaaaacytttcctcoaaagtaaccttcattatgacgaagaatacctctcatcagctgttg 120  
|||||  
61 CAGGAAGAAGTTTCCTCCAAAGTACCTTCATTATGACGAAGAAGACTCTCATCAGCTGTG 120

Ddb  
Qy

121 ttgtgaaaaatgctctctcggtaccttacctaataaaccaaacactacagcaaatggaaagcc 180  
|||||  
121 TTGTGACAAATGTCCTCTGGTACCTTAACAACACACTGTACAGCAAAAGTGGAAGACC 180

Ddb  
Qy

181 gttgtgccccttggcctgacactactacacagacagctggcacacagtgacgagtgt 240  
|||||  
181 GTGTGGGCCCTTGGCCTGACCACTACTACACAGACAGCTGGCACACCAAGTGACGAGTGT 240

Ddb  
Qy

241 ctatactgcagccccgctgtgcaagagtgtagtacgtccaagcaggagtgcaatcgcaacc 300  
|||||  
241 CTATACTGCAGCCCCGCTGTGAAGGAGTGCAGTAGTACGTCAAGCAGGAGTCCAATCGGACC 300

Ddb  
Qy

301 cacacccgctgtgcgaatgcaaggaaggcgctaccttgatatagattctgtctgaaa 360  
|||||  
301 CACAACCGCTGTGGAATGCAAGGAAGGGCGCTACCTTGATAGATAGATTCTGCTTGAAA 360

Ddb  
Qy

361 cataggagctgccctcctgatttgagttggtgcaagctggaaccccgagcgcaataaca 420  
|||||  
361 CATAGGAGCTGCCCTCCTGGATTGTGGAGTGGTGCAAGCTGGAACCCAGAGCGCAAATACA 420

Ddb  
Qy

421 gtttgcaaaagatgtccagatgggttctctctcaaatgagaagtcatctaaagcacccctgt 480  
|||||  
421 GTTTGCAAAAGATGTCCAGATGGGTCTCTCTCAAAATGAGAGCTCATCTAAAGCACCCCTGT 480

Ddb  
Qy

481 agaaaaacacaaaattgcagtgctcttggtctcctgctaactccagaaggaaatgcaaca 540  
|||||  
481 AGAAAAACACAAAATTGTCAGTGTCTTTGGTCTCTCTGCTAACCTCAGAAAGGAANTGCAACA 540

RESULT 9  
ID T33162 standard; DNA; 1206 BP.  
AC T33162;  
DT 22-APR-1997 (first entry)  
DE Mutated OCIF, OCIF-C20S, coding sequence.  
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;

Db 541 caagcaacatattgtccggaacagtgaaatcaactcaaaatgtggaatagattacc 600  
 Qy 541 CACGACAAATATGTTCCGGAACAGTGAATCAACTCAAAATGTGGATAGATTACC 600  
 Db 601 ctgagtgaggagcattcttcacggttggctgtcctacaaagttaacgcttaactggctt 660  
 Qy 601 CTGAGTGAGGAGGAGGATCTTTCAGGTTTGTCTTCCACAAAGTTTACGGCTAACTGGCTT 660  
 Db 661 agtgcctgtgtagacaatttgcctggcaccacaaagtaaacacagagagtgtagaggata 720  
 Qy 661 AGTGCTGTGTAGACAAATTTGCTGCGACCAAGTAAGTAAACGACAGAGTGTAGAGGATA 720  
 Db 721 aaacgcaacacagctcacaaagacagatttccagctgctgaagttaattggaacatcaa 780  
 Qy 721 AAACGGCAACACAGCTCACAAAGAACAGACTTCCAGCTGCTGAAGTTATGGAACATCAA 780  
 Db 781 aacaagacacaaagatagatcaagaagatcatccaaagatatgaactctgtgaaacagc 840  
 Qy 781 AACAAGACCAAGATATAGTCAGAAGATCATCCAAAGATATTGACCTCTGTGAAACACAGC 840  
 Db 841 gtgcagcgcacattggacatgctaactca 871  
 Qy 841 GTGCAGCGGCACATTGGACATGCTAACCTCA 871

RESULT 10  
 ID T33163 standard; DNA: 1206 BP.

AC T33163;  
 DT 22-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-C21S, coding sequence.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 OS Osteoporosis; ss.

FT Key Location/Qualifiers

FT sig\_peptide 1..63

FT /\*tag= a

FT mat\_peptide 64..1203

FT /\*tag= b

FT /product= OCIF-C21S

PN W09626217-AL.

PD 29-AUG-1996.

PF 20-FEB-1996; J00374.

PR 20-FEB-1995; JP-054977.

PR 21-JUL-1995; JP-207508.

PA (SNOW) SNOW BRAND MILK PROD CO LTD.

Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;

Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;

WPI: 96-402320/40.

DR P-PSDB; R99933.

PT DNA encoding osteoclastogenesis inhibitory factor protein - useful

PT for bone resorption control, esp. treatment of osteoporosis

PS Claim 33; Page 134-135; 183pp; Japanese.

CC This sequence encodes a mutated version of the full length

CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This

CC sequence encodes OCIF-C21S in which the 21st Cys residue in the mature

CC OCIF protein is substituted by Ser. The OCIF of the invention has a

CC molecular weight by SDS-PAGE of 60 kD under reducing conditions and

CC 120 kD under non-reducing conditions. The protein is adsorbed onto

CC cation-exchangers or heparin and its activity is lowered after 10 mins

CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90

CC deg.C. OCIF is useful in the control of bone resorption and therefore

CC in the treatment and prevention of disorders of bone resorption, e.g.

CC osteoporosis.

CC Sequence 1206 BP; 389 A; 286 C; 267 G; 264 T;

Query Match 79.4%; Score 865; DB 27; Length 1206;

Best Local Similarity 99.7%; Pred. No. 0.00e+00;

Matches 868; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 atgaacaaactgtgtgctgcgcgctgtgttcttgacatctccattgaattgagacacc 60

Qy 1 ATGAACAAGTGTGCTGCTGCTGCGGCTGTGTCTTGACATCTCCATTAAGTGACACC 60

Db 61 caggaacagtttctccaaagtagtaaccttcaattatgaagaagaacacctctcatcagctgttg 120  
 Qy 61 CAGGAAACGTTTCTCTCAAAGTACCTTCAATTATGACGAAGAAACCTCTCATCAGCTGTTG 120  
 Db 121 tgtacaaaatgtctctctgttacctacctaataaacaacactgtacagcaaatgtggaagacc 180  
 Qy 121 TGTGACAAATGTCTCTCTGTACTACTACCTAAAACAACACTGTACAGCAAGTGGAAAGACC 180  
 Db 181 gtgtgcctcccttgcctgaccactactacacagacagctggcacaccactcagcagtggt 240  
 Qy 181 GTGTGCGCCCTTGGCCCTGACCCTACTACTACACAGACAGCTGGCACACCTGACGAGTGT 240  
 Db 241 ctatctgcagccctgtgtgcaaggagctgcagtagctcaagcagagagtgcaatgcacc 300  
 Qy 241 CTATCTGTCAGCCCTGTGTGCAAGTGGAGAGCTGCAGTCAAGTCAAGCAGGAGTCCAATCCAGC 300  
 Db 301 cacaacgcgtgtgcgaatgcgaaggaggcgctactcttgagatagagttctgcttgaag 360  
 Qy 301 CACAACGCGCTGTGCGAATGCAAGGAAGGCGCTACTCTTGAGATAGAGTCTTGCTTGAAA 360  
 Db 361 cataggagctgcctcctcctgatttgagtggtgcaagctggcaacccacagcgaaataca 420  
 Qy 361 CATAGGAGCTGCCCTCTCTCGAATTTGGAGTGGTCAAGCTGGAACCCACAGCGAAATACA 420  
 Db 421 gtttgcaaaagatgtccagatgggttcttctcaaatgagacgtctaatcctaaagcacccgtg 480  
 Qy 421 GTTTGCAAAAGATGTCCAGATGGGTCTTCTCAATAGACCTCATCTAAAGCACCCCTGT 480  
 Db 481 agaaacacacaaatgtcagtgctcttctcctcctgtaactcagaagaagaaatgaaca 540  
 Qy 481 AGAAAAACACACAATTTGCAGTGTCTTTGGTCTCTCTTAACTCAGAAAGGAATCAACA 540  
 Db 541 cagcaacacatgttccgcgaacagtgaaatcaactcaaaatgtggatagatgttacc 600  
 Qy 541 CAGCACAAATATGTTCCGGAACACAGTGAATCAACTCAAAAAATGGAATAGATGTACC 600  
 Db 601 ctgtgtgaggaggcatcttccaggtttgtgttctcacaagtttacccctaaactggtt 660  
 Qy 601 CTGTGTGAGGAGGATCTTCTCAGGTTTGTGTCTCTCAAAAGTTTACGCCCTAACTGGCTT 660  
 Db 661 agtgcctgtgtgagacaaattgctgcgcacaaagttaacacgagagagtgtagagaggata 720  
 Qy 661 AGTGCTCTGTGTAGACAATTTGCTGCTGCGACCAAGTAAACGACAGAGTGTAGAGAGGATA 720  
 Db 721 aaacggcaacacagctcacaagaacagacttccagctgcgtgaaattatggaacatcaa 780  
 Qy 721 AAACGGCAACACAGCTCACAGAACAGACTTTCAGCTGCTGAAGTTATGGAACATCAA 780  
 Db 781 aacaagacacaaagatatagtcagaagatcatccaagatttgacctcagtgaaacagc 840  
 Qy 781 AACAAGACCAAGATATAGTCAAGAAGATCATCCAAGATATTGACCTCTGTGAAACACAGC 840  
 Db 841 gtgcagcgcacattggacatgctaactca 871  
 Qy 841 GTGCAGCGGCACATTGGACATGCTAACCTCA 871

RESULT 11

ID T33161 standard; DNA: 1206 BP.

AC T33161;

DT 22-APR-1997 (first entry)

DE Mutated OCIF, OCIF-C19S, coding sequence.

KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;

OS Osteoporosis; ss.

FT Key Location/Qualifiers

FT sig\_peptide 1..63

FT /\*tag= a

FT mat\_peptide 64..1203

FT /\*tag= b

FT /product= OCIF-C19S

PN W09626217-AL.

Query Match	74.9%	Score 816;	DB 27;	Length 984;
Best Local Similarity	99.9%	Pred. No. 0.00e+00;		
Matches 817;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
Db	1	atgaacaactgtctgtcgcgcctcgtgttctcgacatctccattaaatggaccacc	60	
QY	1	ATGAACAAGTTGCTGTGCTGGCGCTCGTGTTCGGACATCTCCATTAAATGGACCACC	60	
Db	61	caggaaacgctttctccaaagtaccttcattatgacgaagaacacctctcatcagctgttg	120	
QY	61	CAGGAACCGTTTCTCCAAAGTACCTTCATTATGACGAAGAAACCTCTCATCAGCTGTTC	120	
Db	121	tgtgacaatatgtcctcctgtgtacactaaacacacactgtacagcaaatggaagacc	180	
QY	121	TGTGACAAATGTCTCTCTGTGTACCTAACTAAACAAACACTGTACAGCAAGTGGAGACC	180	
Db	181	atgtscgcgcctctgccttgaccactactacacagacagctggcacaccagtgcagagttg	240	

QY 181 GTGTGCGCCCTTGCCCTGACCACTACTACACAGAGCTGGCACACCACTGACGAGTGT 240  
 Db 241 ctatactgcagcccggtgtcaaggagctcagtagctcaagcagagtgcaatgcacc 300  
 QY 241 CTATACCTGACGCCCGGTGTCAGAGGAGTGCAGTAGCTCAAGCAGAGTGCAATGCACC 300  
 Db 301 caaacccgctgtgcgaatcaaggagggcgctactctgagatagattctgcttgaa 360  
 QY 301 CACAACCGCTGTGCGAATCAAGGAAGGCGCTACCTTGAGATAGATTCTGCTGAAA 360  
 Db 361 catagagctgcctcctgattgagtggtcagagctggaacccagagagaataca 420  
 QY 361 CATAGGAGCTGCCCTCTGATTTGAGTGGTGCAGAGTGGAAACCCAGCGAATAACA 420  
 Db 421 gttgcaaaagatgtccagatgggtctctcaaatgagacgtcatctaaagcaccctgt 480  
 QY 421 GTTTGCAAAAGATGTCCAGATGGTCTTCTCAAAATGACAGCTCACTTAAACGACCCCTGT 480  
 Db 481 agaaacacacaaattgcagtgcttctgtctcctgctactcaagaaagaaatgcaaca 540  
 QY 481 AGAAACACACAAATTCAGTGTCTTTGGTCTCTGCTTAACTCAAGAAAGAAATGCAACA 540  
 Db 541 cagcaacaatatgttccggaacagtgatcaactcaaaatggaatgtggaatattacc 600  
 QY 541 CACGACAACATATGTTCCGGAACACAGTGAATCAACTCAAAATGTGGAATAGATTACC 600  
 Db 601 ctgtgagaggagcatctctcaggtgtgtctcctacaaagtctacgctaaactggctt 660  
 QY 601 CTGTGTGAGGAGGAGCATCTCTCAGTTGTGTCTTCTTACAAAGTTTACGCTAACTGGCTT 660  
 Db 661 agtgcttctgtgacaaattgctgcacccaagtaaacacagagagtgagagaggata 720  
 QY 661 AGTGCTCTGGTAGACAAATTTGCTGCGACCAAGTAACGACGAGAGTGTAGAGAGGATA 720  
 Db 721 aaacgcaacacagctcacaagaacagactttccagctgctgaagtattggaacatcaa 780  
 QY 721 AAACGCAACACAGCTCACAAGAACAGACTTCCAGCTGCTCAAGTTATGGAACATCAA 780  
 Db 781 aacaaagacagatatagtcagaagatcatccaaga 818  
 QY 781 AACNAAGCCACAGATATAGTCAAGAAGATCATCCAAGA 818

## RESULT 13

ID T33174 standard; DNA; 819 BP.

T33174.

22-APR-1997 (first entry)

Mutated OCIF, OCIF-CDD2, coding sequence.

KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;

OS Osteoporosis; ss.

FT key

FT sig\_peptide Location/Qualifiers

FT /\*tag a

FT mat\_peptide 64..816

FT /\*tag b

FT /product= OCIF-CDD2

PN W09626217-A1.

PD 29-AUG-1996.

PF 20-FEB-1996; J00374.

PR 20-FEB-1995; JP-054977.

PR 21-JUL-1995; JP-207508.

PA (SNOW) SNOW BRAND MILK PROD CO LTD.

PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;

PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;

DR WPI; 96-402320/40.

DR P-PSDB; R39944.

PT DNA encoding osteoclastogenesis inhibitory factor protein - useful

PT for bone resorption control, esp. treatment of osteoporosis

PS Claim 66; Page 145; 183pp; Japanese.

CC This sequence encodes a mutated version of the full length

CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This

## RESULT 14

ID T33166 standard; DNA; 1083 BP.

CC sequence encodes OCIF-CDD2 in which amino acids 252-380 of the mature  
 CC protein have been deleted. The OCIF of the invention has a molecular  
 CC weight by SDS-PAGE of 60 kD under reducing conditions and 120 kD under  
 CC non-reducing conditions. The protein is adsorbed onto cation-exchangers  
 CC or heparin and its activity is lowered after 10 mins at 70 deg.C or 30  
 CC mins at 56 deg.C, and is lost after 10 mins at 90 deg.C. OCIF is useful  
 CC in the control of bone resorption and therefore in the treatment and  
 CC prevention of disorders of bone resorption, e.g. osteoporosis.  
 SQ Sequence 819 BP; 251 A; 198 C; 188 G; 182 T;

Query Match 74.7%; Score 814; DB 27; Length 819;

Best Local Similarity 99.9%; Pred. No. 0.00e+00;

Matches 815; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 atgaacaactgtgtgctgcgcgcgtggttctctggacatctccatttaagtggaccac 60  
 QY 1 ATGAACAAGTGTGTGCTGCGCGCTCGTGTCTTGACATCTCCATTAAAGTGGACACC 60  
 Db 61 caggaaacgtttctccaaagtacacctcattatgacgaagaacacctctcatcagctgtg 120  
 QY 61 CAGGAAACGTTTCTCCAAAGTACCTTCATTATGACGAAGAACCCTCTCATCAGCTGTG 120  
 Db 121 tgtgacaaatgtcctcctcctgctacacctcctaaacacactgtacagcaagtggagacc 180  
 QY 121 TGTGACAAATGTCTCTCTGCTACCTTACCTAAAACAACACTGTACAGCAAGTGGAAAGACC 180  
 Db 181 gtgtgcgcctctgctcctaccactactacacagacagctggcacaccagtgacagagt 240  
 QY 181 GTGTGCGCCCTTGCCTGACCACTACTACACAGACAGCTGGCACACCAGTGCAGAGTGT 240  
 Db 241 ctatactgcagcccgctgtgcgaaggctgcagtagctcgaagcagagtgcaatgcacc 300  
 QY 241 CTATACTGCAGCCCGCTGTGCAAGGAGCTGCAGTACGTCAAGCAGGAGTGCATCGCACC 300  
 Db 301 cacaacgcgtgtgcgaatgcgaaggcgcctcactcctcctcctcctcctcctcctcct 360  
 QY 301 CACAACGCGTGTGCGAATGCAGAGGAGGCGCTACCTTTGAGATAGAGTCTTGCCTGAAA 360  
 Db 361 catagagctgcctcctcctgattggagtggtgcaagctggacccagagcagaataca 420  
 QY 361 CATAGGAGCTGCCCTCTCTGGATTGGAGTGGTGAAGCTGGAACCCAGAGCGAATACA 420  
 Db 421 gtttgcacaaagatgtccagatgggttctctcctcctcctcctcctcctcctcctcct 480  
 QY 421 GTTTGCAAAAGATGTCCAGATGGTCTTCTCAAAATGAGACGTCACTTAAAGCACCCCTGT 480  
 Db 481 agaaacacacacaaattgcagtgcttctgtctcctgctcctcctcctcctcctcctcct 540  
 QY 481 AGAAACACACAAATTCAGTGTCTTTGGTCTCTCTGCTTAACCTCAGAAAGGAATGCAACA 540  
 Db 541 cagcaacaatatgttcccggaacagtggaatcaactcacaatgtggaatagattacc 600  
 QY 541 CACGACAACATATGTTCCGGAACACAGTGAATCAACTCAAAATGTGGAATAGATTACC 600  
 Db 601 ctgtgtgagagagcatctcctcaggtttctgttctcctcctcctcctcctcctcctcct 660  
 QY 601 CTGTGTGAGGAGGAGCATCTCTCAGTTTGTCTCTTCTTCTTCTTCTTCTTCTTCTTCTT 660  
 Db 661 agtgctgtgtgtagacaatttgctgcgcaccacaaagttaacgcagagtgtagagaggata 720  
 QY 661 AGTGCTTGTGTAGACAATTTGCTGGCACCACCAAGTAAACGACAGAGTGTAGAGAGGATA 720  
 Db 721 aaacgcaacacagctcacaagaacagactttcagctgtcgtgaagttatggaacatcaa 780  
 QY 721 AAACGCAACACAGCTCACAAGAACAGACTTCCAGCTGCTCAAGTTATGGAACATCAA 780  
 Db 781 aacaagacccaagatatagtcagaagatcatcca 816  
 QY 781 AACAAAGACCAAGATATAGTCAAGAGATCATCCA 816



|||||  
QY 181 GTGTCGCCCTTGCCCTGACCACTACTACACAGACAGCTGGCACACCACTGACGAGTGT 240  
Db 241 ctatactcagcccgctgtgcaagagctgcagtacgtcaagcagagtgcaatcgacc 300  
QY 241 CTATACTCAGCCCGTGTGCAAGAGCTGCAGTACGTCAAGCAGGAGTGCAATCGCACC 300  
Db 301 cacaaccgctgtgcgaatgcaaggaaggcgctacaccttgagatagagttctgttgaaa 360  
QY 301 CACAACCGCTGTGGAATGCAAGGAAGGGCGCTACCTTGAGATAGAGTCTGCTTGAAA 360  
Db 361 cataagagctgccctcctctggatttggagtggtgcaagctggaaccccgagagcgaaataca 420  
QY 361 CATAGGAGCTGCCCTCCCTGGATTGGAGTGGTGCAAGCTGGAAACCCCGAGAGCGAAATACA 420  
Db 421 gtttgcaaaagatgtccagatgggttctctcaaatgagacgtcatctaaagcacccctgt 480  
QY 421 GTTTGCAAAAGATGTCCAGATGGGTTCTTCTCAATGAGACGTCTAAGCACCCCTGT 480  
Db 481 agaaaacacacaaattgcagtgtcttgggtctctctgtaactcagaaggaagaaatgcaaca 540  
QY 481 AGAAAACACACAAATTGCAGTGTCTTTGGTCTCCTGCTAACTCAGAAAGGAAATGCAACA 540  
Db 541 cagcaacacatatctccggaaacagtgaaatcaactcaaaaatgtggaatagatatggac 600  
QY 541 CACGCAACATATGTTCGGAAACAGTGAATCAACTCAAAATGTGGAATAGATGTATACC 600  
Db 601 ctctgtga 608  
QY 601 CTGTGTGA 608

Search completed: Tue Dec 2 16:40:26 1997  
Job time : 262 secs.

\*\*\*\*\*

(TM)

\*\*\*\*\*

Release 2.1b John F. Collins, Biocomputing Research Unit.  
 Copyright (C) 1993, 1994, 1995 University of Edinburgh, U.K.  
 Distribution rights by Intelligent, Inc.

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Dec 3 10:06:20 1997; MasPar time 2.51 seconds  
 29.342 Million cell updates/sec

ular output not generated.

File: >US-08-915-004-1  
 Description: (1-6) from US08915004.pep  
 Perfect Score: 47  
 Sequence: 1 XYHFPK 6

Scoring table:  
 PAM 150  
 Gap 15

Searched: 101610 seqs, 12294212 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: a-geneseq28  
 1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
 14:part14 15:part15 16:part16 17:part17 18:part18  
 19:part19 20:part20 21:part21

Statistics: Mean 14.688; Variance 38.315; scale 0.383

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

Hit No.	Score	Match	Length	ID	Description	Pred. No.
1	47	100.0	6	20	R99921 Osteoclastogenesis in	2.01e+01
2	47	100.0	326	20	R99940 Mutated OCIF, OCIF-DD	2.01e+01
3	47	100.0	327	20	R99941 Mutated OCIF, OCIF-DD	2.01e+01
4	47	100.0	359	20	R99939 Mutated OCIF, OCIF-DC	2.01e+01
5	47	100.0	359	20	R99937 Mutated OCIF, OCIF-DC	2.01e+01
6	47	100.0	360	20	R99938 Mutated OCIF, OCIF-DC	2.01e+01
7	47	100.0	360	20	R99936 Mutated OCIF, OCIF-DC	2.01e+01
8	47	100.0	380	20	R99924 Mature osteoclastogen	2.01e+01
9	47	100.0	393	20	R99948 Mutated OCIF, OCIF-CB	2.01e+01
10	47	100.0	399	20	R99942 Mutated OCIF, OCIF-CL	2.01e+01
11	47	100.0	401	20	R99935 Mutated OCIF, OCIF-C2	2.01e+01
12	47	100.0	401	20	R99932 Mutated OCIF, OCIF-C2	2.01e+01
13	47	100.0	401	20	R99931 Mutated OCIF, OCIF-C1	2.01e+01
14	47	100.0	401	20	R99934 Mutated OCIF, OCIF-C2	2.01e+01
15	47	100.0	401	20	R99933 Mutated OCIF, OCIF-C2	2.01e+01
16	47	100.0	401	20	R99925 Full length osteoclas	2.01e+01
17	43	91.5	505	1	P82665 Sequence encoded by p	6.35e+01
18	42	89.4	1137	15	R90766 Tumour suppressor pro	8.43e+01
19	40	85.1	188	12	R66461 AF-17 protein, N-term	1.47e+02
20	40	85.1	255	4	R23005 Protein transcribed f	1.47e+02

21	40	85.1	390	20	R99357 Human tumour necrosis	1.47e+02
22	40	85.1	422	19	W00729 Daunorubicin 14-hydro	1.47e+02
23	40	85.1	527	18	R91277 Exophiala spinifera f	1.47e+02
24	40	85.1	934	3	R14402 Soluble insulin-like	1.47e+02
25	40	85.1	934	3	R15048 Soluble human IGF-I r	1.47e+02
26	40	85.1	948	3	R14403 Extracellular domain	1.47e+02
27	40	85.1	948	3	R15049 Hybrid human insulin-	1.47e+02
28	40	85.1	1018	15	R87028 Human contactin	1.47e+02
29	40	85.1	1018	11	R63759 Human contactin (EMBL	1.47e+02
30	40	85.1	1093	12	R66460 AF-17 protein.	1.47e+02
31	40	85.1	1337	11	R63125 IGF-I receptor 957A v	1.47e+02
32	40	85.1	1337	11	R63126 IGF-I receptor 1003A	1.47e+02
33	40	85.1	1337	11	R63124 IGF-I receptor 950X v	1.47e+02
34	40	85.1	1337	11	R63123 IGF-I receptor 943A v	1.47e+02
35	40	85.1	1367	11	R60795 Human IGF-1 receptor.	1.47e+02
36	40	85.1	1367	16	R91429 Human type I insulin-	1.47e+02
37	40	85.1	1367	17	R95244 IGF-1 receptor.	1.47e+02
38	40	85.1	1367	11	R63122 IGF-1 receptor.	1.47e+02
39	40	85.1	1370	16	R91430 Rat type I insulin-li	1.47e+02
40	39	83.0	214	19	W06503 Htm4 protein.	1.94e+02
41	39	83.0	821	18	R99579 Calpain large subunit	1.94e+02
42	38	80.9	248	21	W20433 H. pylori secreted or	2.54e+02
43	38	80.9	327	15	R77379 Alteromonas nandedal	2.54e+02
44	38	80.9	566	4	R20181 Sequence encoded by 6	2.54e+02
45	38	80.9	590	1	R07505 Fragment of Plasmodiu	2.54e+02

## ALIGNMENTS

RESULT 1  
 ID R99921 standard; peptide; 6 AA.  
 AC R99921;  
 DT 22-APR-1997 (first entry)  
 DE Osteoclastogenesis inhibitory factor peptide fragment #1.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 OS osteoporosis.  
 KW Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Misc\_difference 1  
 FT /note="Any amino acid"  
 PN W09626217-AL.  
 PD 29-AUG-1996.  
 PR 20-FEB-1996; JP-054977.  
 PR 20-FEB-1995; JP-207508.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 PI WPI; 96-402320/40.  
 DR DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control. esp. treatment of osteoporosis  
 PS Claim 1; Page 61; 183pp; Japanese.  
 CC The sequences given in R99921-23 and R99926 represent fragments of  
 CC the osteoclastogenesis inhibitory factor (OCIF) of the invention.  
 CC The OCIF has a molecular weight by SDS-PAGE of 60 kD under reducing  
 CC conditions and 120 kD under non-reducing conditions. The protein is  
 CC adsorbed onto cation-exchangers or heparin and its activity is  
 CC lowered after 10 mins at 70 deg.C or 30 mins at 56 deg.C, and is  
 CC lost after 10 mins at 90 deg.C. OCIF is useful in the control of  
 CC bone resorption and therefore in the treatment and prevention of  
 CC disorders of bone resorption, e.g. osteoporosis.  
 SQ Sequence 6 AA;

Query Match 100.0%; Score 47; DB 20; Length 6;  
 Best Local Similarity 100.0%; Pred. No. 2.01e+01;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2 yhfpk 6

Qy 2 YHFPK 6

RESULT 2



ID R99940 standard; Protein; 326 AA.  
AC R99940; 23-APR-1997 (first entry)  
DE Mutated OCIF, OCIF-DDD1.  
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
OS osteoporosis.  
KW Synthetic.  
FH Key Location/Qualifiers  
FT Peptide 1..21  
FT /note= "Signal peptide"  
FT Protein 22..326  
FT /note= "Mature OCIF-DDD1"  
FT Misc.difference 198..199  
FT /note= "Position of deletion, delta 178-252"  
PN WO9626217-A1.  
PD 29-AUG-1996.  
PF 20-FEB-1996; J00374.  
PR 20-FEB-1995; JP-054977.  
PR 21-JUL-1995; JP-207508.  
PA (SNOW ) SNOW BRAND MILK PROD CO LTD.  
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
DR WPI: 96-402320/40.  
DR N-PSDB: T33170.  
PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
PT for bone resorption control, esp. treatment of osteoporosis  
PS Claim 36; Page 113-114; 183pp; Japanese.  
CC This sequence represents a mutated version of the full length  
CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
CC sequence represents OCIF-DDD1 in which amino acids 178-252 of the  
CC mature OCIF protein are deleted. The OCIF of the invention  
CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
CC and 120 kD under non-reducing conditions. The protein is adsorbed onto  
CC cation-exchangers or heparin and its activity is lowered after 10 mins  
CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
CC deg.C. OCIF is useful in the control of bone resorption and therefore  
CC in the treatment and prevention of disorders of bone resorption, e.g.  
CC osteoporosis.  
SQ Sequence 326 AA;  
Query Match 100.0%; Score 47; DB 20; Length 326;  
Best Local Similarity 100.0%; Pred. No. 2.01e+01;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 280 yhfpk 284  
|||||  
2 YHFPK 6  
RESULT 3  
ID R99941 standard; Protein; 327 AA.  
AC R99941; 23-APR-1997 (first entry)  
DE Mutated OCIF, OCIF-DDD2.  
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
OS osteoporosis.  
KW Synthetic.  
FH Key Location/Qualifiers  
FT Peptide 1..21  
FT /note= "Signal peptide"  
FT Protein 22..327  
FT /note= "Mature OCIF-DDD2"  
FT Misc.difference 273..274  
FT /note= "Position of deletion, delta 253-326"  
PN WO9626217-A1.  
PD 29-AUG-1996.  
PF 20-FEB-1996; J00374.  
PR 20-FEB-1995; JP-054977.  
PR 21-JUL-1995; JP-207508.  
PA (SNOW ) SNOW BRAND MILK PROD CO LTD.  
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
DR WPI: 96-402320/40.

DR N-PSDB; T33171.  
PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
PT for bone resorption control, esp. treatment of osteoporosis  
PS Claim 59; Page 115-116; 183pp; Japanese.  
CC This sequence represents a mutated version of the full length  
CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
CC sequence represents OCIF-DDD2 in which amino acids 253-326 of the  
CC mature OCIF protein are deleted. The OCIF of the invention  
CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
CC and 120 kD under non-reducing conditions. The protein is adsorbed onto  
CC cation-exchangers or heparin and its activity is lowered after 10 mins  
CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
CC deg.C. OCIF is useful in the control of bone resorption and therefore  
CC in the treatment and prevention of disorders of bone resorption, e.g.  
CC osteoporosis.  
SQ Sequence 327 AA;  
Query Match 100.0%; Score 47; DB 20; Length 327;  
Best Local Similarity 100.0%; Pred. No. 2.01e+01;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 281 yhfpk 285  
|||||  
2 YHFPK 6  
RESULT 4  
ID R99939 standard; Protein; 359 AA.  
AC R99939; 23-APR-1997 (first entry)  
DE Mutated OCIF, OCIF-DCR4.  
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
OS osteoporosis.  
KW Synthetic.  
FH Key Location/Qualifiers  
FT Peptide 1..21  
FT /note= "Signal peptide"  
FT Protein 22..359  
FT /note= "Mature OCIF-DCR4"  
FT Misc.difference 143..144  
FT /note= "Position of deletion, delta 123-164"  
PN WO9626217-A1.  
PD 29-AUG-1996.  
PF 20-FEB-1996; J00374.  
PR 20-FEB-1995; JP-054977.  
PR 21-JUL-1995; JP-207508.  
PA (SNOW ) SNOW BRAND MILK PROD CO LTD.  
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
DR WPI: 96-402320/40.  
DR N-PSDB; T33169.  
PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
PT for bone resorption control, esp. treatment of osteoporosis  
PS Claim 53; Page 111-113; 183pp; Japanese.  
CC This sequence represents a mutated version of the full length  
CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
CC sequence represents OCIF-DCR4 in which amino acids 123-164 of the  
CC mature OCIF protein are deleted. The OCIF of the invention  
CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
CC and 120 kD under non-reducing conditions. The protein is adsorbed onto  
CC cation-exchangers or heparin and its activity is lowered after 10 mins  
CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
CC deg.C. OCIF is useful in the control of bone resorption and therefore  
CC in the treatment and prevention of disorders of bone resorption, e.g.  
CC osteoporosis.  
SQ Sequence 359 AA;  
Query Match 100.0%; Score 47; DB 20; Length 359;  
Best Local Similarity 100.0%; Pred. No. 2.01e+01;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 313 yhfpk 317  
|||||

Qy 2 YHFPK 6

RESULT 5  
 ID R99937 standard; Protein; 359 AA.  
 AC R99937; 23-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-DCR2.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 KW osteoporosis.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /note= "Signal peptide"  
 FT Protein 22..359  
 FT /note= "Mature OCIF-DCR2"  
 FT Misc.difference 63..64  
 FT /note= "Position of deletion, delta 43-84"  
 PN W09626217-Al.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW ) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR N-PSDB; T33167.  
 PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 47; Page 107-109; 183pp; Japanese.  
 CC This sequence represents a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC sequence represents OCIF-DCR2 in which amino acids 43-84 of the  
 CC mature OCIF protein are deleted. The OCIF of the invention  
 CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
 CC and 120 kD under non-reducing conditions. The protein is adsorbed onto  
 CC cation-exchangers or heparin and its activity is lowered after 10 mins  
 CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
 CC deg.C. OCIF is useful in the control of bone resorption and therefore  
 CC in the treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 SQ Sequence 359 AA;

Query Match 100.0%; Score 47; DB 20; Length 359;  
 Best Local Similarity 100.0%; Pred. No. 2.01e+01;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 313 yhfpk 317  
 |||||  
 2 YHFPK 6

RESULT 6  
 ID R99938 standard; Protein; 360 AA.  
 AC R99938;  
 DE 23-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-DCR3.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 KW osteoporosis.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /note= "Signal peptide"  
 FT Protein 22..360  
 FT /note= "Mature OCIF-DCR3"  
 FT Misc.difference 105..106  
 FT /note= "Position of deletion, delta 85-122"  
 PN W09626217-Al.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; JP-054977.  
 PR 21-JUL-1995; JP-207508.

PA (SNOW ) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR N-PSDB; T33168.  
 PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 50; Page 109-111; 183pp; Japanese.  
 CC This sequence represents a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC sequence represents OCIF-DCR3 in which amino acids 85-122 of the  
 CC mature OCIF protein are deleted. The OCIF of the invention  
 CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
 CC and 120 kD under non-reducing conditions. The protein is adsorbed onto  
 CC cation-exchangers or heparin and its activity is lowered after 10 mins  
 CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
 CC deg.C. OCIF is useful in the control of bone resorption and therefore  
 CC in the treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 SQ Sequence 360 AA;

Query Match 100.0%; Score 47; DB 20; Length 360;  
 Best Local Similarity 100.0%; Pred. No. 2.01e+01;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 317 yhfpk 321  
 |||||  
 2 YHFPK 6

RESULT 7  
 ID R99936 standard; Protein; 360 AA.  
 AC R99936;  
 DE 23-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-DCR1.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 KW osteoporosis.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /note= "Signal peptide"  
 FT Protein 22..360  
 FT /note= "Mature OCIF-DCR1"  
 FT Misc.difference 22..23  
 FT /note= "Position of deletion, delta 2-42"  
 PN W09626217-Al.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW ) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR N-PSDB; T33166.  
 PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 44; Page 105-107; 183pp; Japanese.  
 CC This sequence represents a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC sequence represents OCIF-DCR1 in which amino acids 2-42 of the  
 CC mature OCIF protein are deleted. The OCIF of the invention  
 CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
 CC and 120 kD under non-reducing conditions. The protein is adsorbed onto  
 CC cation-exchangers or heparin and its activity is lowered after 10 mins  
 CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
 CC deg.C. OCIF is useful in the control of bone resorption and therefore  
 CC in the treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 SQ Sequence 360 AA;

Query Match 100.0%; Score 47; DB 20; Length 360;  
 Best Local Similarity 100.0%; Pred. No. 2.01e+01;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 314 yhfpk 318  
|||||

Qy 2 YHFPK 6

RESULT 8

ID R9924 standard; Protein; 380 AA.

AC R9924;

DT 22-APR-1997 (first entry)

DE Mature osteoclastogenesis inhibitory factor.

KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;

KW Osteoporosis.

OS Homo sapiens.

PN WO9626217-A1.

PD 29-AUG-1996.

PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;

PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;

DR N-PSDB; T36685.

PT DNA encoding osteoclastogenesis inhibitory factor protein - useful

PT for bone resorption control, esp. treatment of osteoporosis

PS Claim 6; Page 62-64; 183pp; Japanese.

CC This sequence represents the mature osteoclastogenesis inhibitory

CC factor (OCIF) of the invention. The OCIF has a molecular weight by

CC SDS-PAGE of 60 kD under reducing conditions and 120 kD under non-

CC reducing conditions. The protein is adsorbed onto cation-exchangers

CC or heparin and its activity is lowered after 10 mins at 70 deg.C or

CC 30 mins at 56 deg.C, and is lost after 10 mins at 90 deg.C. OCIF is

CC useful in the control of bone resorption and therefore in the

CC treatment and prevention of disorders of bone resorption, e.g.

CC osteoporosis.

SQ Sequence 380 AA;

Query Match 100.0%; Score 47; DB 20; Length 380;

Best Local Similarity 100.0%; Pred. No. 2.01e-01;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 334 yhfpk 338  
|||||

Qy 2 YHFPK 6

RESULT 9

ID R9948 standard; Protein; 393 AA.

AC R9948;

DT 23-APR-1997 (first entry)

DE Mutated OCIF, OCIF-CBst.

KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;

KW Osteoporosis.

OS Synthetic.

FH Key Location/Qualifiers

FT Peptide 1..21

FT /note= "Signal peptide"

FT Protein 22..393

FT /note= "Mature OCIF-CBst"

PN WO9626217-A1.

PD 29-AUG-1996.

PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;

PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;

DR N-PSDB; T33178.

PT DNA encoding osteoclastogenesis inhibitory factor protein - useful

PT for bone resorption control, esp. treatment of osteoporosis

PS Claim 6; Page 117-119; 183pp; Japanese.

CC This sequence represents a mutated version of the full length

CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This

CC sequence represents OCIF-CL in which amino acids 379-380 of the

CC mature OCIF protein are deleted. The OCIF of the invention

CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions

CC and 120 kD under non-reducing conditions. The protein is adsorbed onto

CC cation-exchangers or heparin and its activity is lowered after 10 mins

CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90

CC deg.C. OCIF is useful in the control of bone resorption and therefore

CC in the treatment and prevention of disorders of bone resorption, e.g.

CC osteoporosis.

SQ Sequence 393 AA;

Query Match 100.0%; Score 47; DB 20; Length 393;

Best Local Similarity 100.0%; Pred. No. 2.01e-01;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

PT DNA encoding osteoclastogenesis inhibitory factor protein - useful

PT for bone resorption control, esp. treatment of osteoporosis

PS Claim 80; Page 126-128; 183pp; Japanese.

CC This sequence represents a mutated version of the full length

CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This

CC sequence represents OCIF-CBst in which Gln 371 is substituted by

CC Leu and amino acids 373-380 of the mature OCIF protein are deleted.

CC These changes are caused by the introduction of a restriction site in

CC the DNA encoding this protein. The OCIF of the invention has a

CC molecular weight by SDS-PAGE of 60 kD under reducing conditions

CC and 120 kD under non-reducing conditions. The protein is adsorbed onto

CC cation-exchangers or heparin and its activity is lowered after 10 mins

CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90

CC deg.C. OCIF is useful in the control of bone resorption and therefore

CC in the treatment and prevention of disorders of bone resorption, e.g.

CC osteoporosis.

SQ Sequence 393 AA;

Query Match 100.0%; Score 47; DB 20; Length 393;

Best Local Similarity 100.0%; Pred. No. 2.01e-01;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 355 yhfpk 359  
|||||

Qy 2 YHFPK 6

RESULT 10

ID R9942 standard; Protein; 399 AA.

AC R9942;

DT 23-APR-1997 (first entry)

DE Mutated OCIF, OCIF-CL.

KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;

KW Osteoporosis.

OS Synthetic.

FH Key Location/Qualifiers

FT Peptide 1..21

FT /note= "Signal peptide"

FT Protein 22..399

FT /note= "Mature OCIF-CL"

PN WO9626217-A1.

PD 29-AUG-1996.

PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;

PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;

DR N-PSDB; T33172.

PT DNA encoding osteoclastogenesis inhibitory factor protein - useful

PT for bone resorption control, esp. treatment of osteoporosis

PS Claim 62; Page 117-119; 183pp; Japanese.

CC This sequence represents a mutated version of the full length

CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This

CC sequence represents OCIF-CL in which amino acids 379-380 of the

CC mature OCIF protein are deleted. The OCIF of the invention

CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions

CC and 120 kD under non-reducing conditions. The protein is adsorbed onto

CC cation-exchangers or heparin and its activity is lowered after 10 mins

CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90

CC deg.C. OCIF is useful in the control of bone resorption and therefore

CC in the treatment and prevention of disorders of bone resorption, e.g.

CC osteoporosis.

SQ Sequence 399 AA;

Query Match 100.0%; Score 47; DB 20; Length 399;

Best Local Similarity 100.0%; Pred. No. 2.01e-01;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 355 yhfpk 359  
|||||

Qy 2 YHFPK 6

```

RESULT 11
ID R99935 standard; Protein: 401 AA.
AC R99935;
DE 22-APR-1997 (first entry)
DT Mutated OCIF, OCIF-C23S.
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;
OS osteoporosis.
FH Key Location/Qualifiers
FT Peptide 1..21
FT /note= "Signal peptide"
FT Protein 22..401
FT /note= "Mature OCIF-C23S"
FT Misc_difference 400
FT /label= C23S
PN WO9626217-A1.
PD 29-AUG-1996. J00374.
PF 20-FEB-1996; JP-054977.
PR 20-FEB-1995; JP-207508.
PS (SNOW ) SNOW BRAND MILK PROD CO LTD.
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;
Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;
WPI: 96-402320/40.
DR N-PSDB; T33165.
DT DNA encoding osteoclastogenesis inhibitory factor protein - useful
PT for bone resorption control, esp. treatment of osteoporosis
PS Claim 41; Page 103-105; 183pp; Japanese.
CC This sequence represents a mutated version of the full length
CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This
CC mature OCIF protein is substituted by Ser. The OCIF of the invention
CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions
CC and 120 kD under non-reducing conditions. The protein is adsorbed onto
CC cation-exchangers or heparin and its activity is lowered after 10 mins
CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90
CC deg.C. OCIF is useful in the control of bone resorption and therefore
CC in the treatment and prevention of disorders of bone resorption, e.g.
CC osteoporosis.
SQ Sequence 401 AA;

Query Match 100.0%; Score 47; DB 20; Length 401;
Best Local Similarity 100.0%; Pred. No. 2.01e-01;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 355 yhfpk 359
QY 2 YHFPK 6

RESULT 12
ID R99932 standard; Protein: 401 AA.
AC R99932;
DE 22-APR-1997 (first entry)
DT Mutated OCIF, OCIF-C20S.
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;
OS osteoporosis.
FH Key Location/Qualifiers
FT Peptide 1..21
FT /note= "Signal peptide"
FT Protein 22..401
FT /note= "Mature OCIF-C20S"
FT Misc_difference 202
FT /label= C20S
PN WO9626217-A1.
PD 29-AUG-1996. J00374.
PF 20-FEB-1996; JP-054977.
PR 20-FEB-1995; JP-207508.
PS (SNOW ) SNOW BRAND MILK PROD CO LTD.
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;
Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;
WPI: 96-402320/40.
DR N-PSDB; T33161.
DT DNA encoding osteoclastogenesis inhibitory factor protein - useful
PT for bone resorption control, esp. treatment of osteoporosis
PS Claim 29; Page 94-96; 183pp; Japanese.
CC This sequence represents a mutated version of the full length
CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This
CC mature OCIF protein is substituted by Ser. The OCIF of the invention
CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions
CC and 120 kD under non-reducing conditions. The protein is adsorbed onto
CC cation-exchangers or heparin and its activity is lowered after 10 mins
CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90
CC deg.C. OCIF is useful in the control of bone resorption and therefore
CC in the treatment and prevention of disorders of bone resorption, e.g.
CC osteoporosis.
SQ Sequence 401 AA;

Query Match 100.0%; Score 47; DB 20; Length 401;
Best Local Similarity 100.0%; Pred. No. 2.01e-01;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 355 yhfpk 359
QY 2 YHFPK 6

```

```

PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;
WPI: 96-402320/40.
DR N-PSDB; T33162.
DT DNA encoding osteoclastogenesis inhibitory factor protein - useful
PT for bone resorption control, esp. treatment of osteoporosis
PS Claim 32; Page 96-98; 183pp; Japanese.
CC This sequence represents a mutated version of the full length
CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This
CC sequence represents OCIF-C20S in which the 20th Cys residue in the
CC mature OCIF protein is substituted by Ser. The OCIF of the invention
CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions
CC and 120 kD under non-reducing conditions. The protein is adsorbed onto
CC cation-exchangers or heparin and its activity is lowered after 10 mins
CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90
CC deg.C. OCIF is useful in the control of bone resorption and therefore
CC in the treatment and prevention of disorders of bone resorption, e.g.
CC osteoporosis.
SQ Sequence 401 AA;

```

```

Query Match 100.0%; Score 47; DB 20; Length 401;
Best Local Similarity 100.0%; Pred. No. 2.01e-01;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 355 yhfpk 359
QY 2 YHFPK 6

```

```

RESULT 13
ID R99931 standard; Protein: 401 AA.
AC R99931;
DE 22-APR-1997 (first entry)
DT Mutated OCIF, OCIF-C19S.
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;
OS osteoporosis.
FH Key Location/Qualifiers
FT Peptide 1..21
FT /note= "Signal peptide"
FT Protein 22..401
FT /note= "Mature OCIF-C19S"
FT Misc_difference 195
FT /label= C19S
PN WO9626217-A1.
PD 29-AUG-1996. J00374.
PF 20-FEB-1996; JP-054977.
PR 20-FEB-1995; JP-207508.
PS (SNOW ) SNOW BRAND MILK PROD CO LTD.
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;
Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;
WPI: 96-402320/40.
DR N-PSDB; T33161.
DT DNA encoding osteoclastogenesis inhibitory factor protein - useful
PT for bone resorption control, esp. treatment of osteoporosis
PS Claim 29; Page 94-96; 183pp; Japanese.
CC This sequence represents a mutated version of the full length
CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This
CC sequence represents OCIF-C19S in which the 19th Cys residue in the
CC mature OCIF protein is substituted by Ser. The OCIF of the invention
CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions
CC and 120 kD under non-reducing conditions. The protein is adsorbed onto
CC cation-exchangers or heparin and its activity is lowered after 10 mins
CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90
CC deg.C. OCIF is useful in the control of bone resorption and therefore
CC in the treatment and prevention of disorders of bone resorption, e.g.
CC osteoporosis.
SQ Sequence 401 AA;

```

```

Query Match 100.0%; Score 47; DB 20; Length 401;
Best Local Similarity 100.0%; Pred. No. 2.01e-01;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Db 355 yhfpk 359  
 |||||  
 QY 2 YHFPK 6

RESULT 14  
 ID R9934 standard; Protein; 401 AA.  
 AC R9934;  
 DT 22-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-C22S  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 OS Osteoporosis.  
 FT Key Synthetic.  
 FT Peptide 1..21  
 FT /note= "Signal peptide"  
 FT Protein 22..401  
 FT /note= "Mature OCIF-C22S"  
 FT Misc.Difference 277  
 FT /label= C22S  
 PN WO9626217-AL.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; J00374.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW ) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR N-PSDB; T33164.

PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 38; Page 100-102; 183pp; Japanese.  
 CC This sequence represents a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC sequence represents OCIF-C22S in which the 22nd Cys residue in the  
 CC mature OCIF protein is substituted by Ser. The OCIF of the invention  
 CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
 CC and 120 kD under non-reducing conditions. The protein is adsorbed onto  
 CC cation-exchangers or heparin and its activity is lowered after 10 mins  
 CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
 CC deg.C. OCIF is useful in the control of bone resorption and therefore  
 CC in the treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 CC Sequence 401 AA;  
 Query Match 100.0%; Score 47; DB 20; Length 401;  
 Best Local Similarity 100.0%; Pred. No. 2.01e+01;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 355 yhfpk 359  
 |||||  
 QY 2 YHFPK 6

RESULT 15  
 ID R9933 standard; Protein; 401 AA.  
 AC R9933;  
 DT 22-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-C21S  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 OS Osteoporosis.  
 FT Key Synthetic.  
 FT Peptide 1..21  
 FT /note= "Signal peptide"  
 FT Protein 22..401  
 FT /note= "Mature OCIF-C21S"  
 FT Misc.Difference 277  
 FT /label= C21S  
 PN WO9626217-AL.  
 PD 29-AUG-1996.

Db 355 yhfpk 359  
 |||||  
 QY 2 YHFPK 6

Query Match 100.0%; Score 47; DB 20; Length 401;  
 Best Local Similarity 100.0%; Pred. No. 2.01e+01;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 15  
 ID R9933 standard; Protein; 401 AA.  
 AC R9933;  
 DT 22-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-C21S  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 OS Osteoporosis.  
 FT Key Synthetic.  
 FT Peptide 1..21  
 FT /note= "Signal peptide"  
 FT Protein 22..401  
 FT /note= "Mature OCIF-C21S"  
 FT Misc.Difference 277  
 FT /label= C21S  
 PN WO9626217-AL.  
 PD 29-AUG-1996.

PF 20-FEB-1996; J00374.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW ) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR N-PSDB; T33163.

PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 35; Page 98-100; 183pp; Japanese.  
 CC This sequence represents a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC sequence represents OCIF-C21S in which the 21st Cys residue in the  
 CC mature OCIF protein is substituted by Ser. The OCIF of the invention  
 CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
 CC and 120 kD under non-reducing conditions. The protein is adsorbed onto  
 CC cation-exchangers or heparin and its activity is lowered after 10 mins  
 CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
 CC deg.C. OCIF is useful in the control of bone resorption and therefore  
 CC in the treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 CC Sequence 401 AA;  
 Query Match 100.0%; Score 47; DB 20; Length 401;  
 Best Local Similarity 100.0%; Pred. No. 2.01e+01;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 355 yhfpk 359  
 |||||  
 QY 2 YHFPK 6

RESULT 16  
 ID R9925 standard; Protein; 401 AA.  
 AC R9925;  
 DT 22-APR-1997 (first entry)  
 DE Full length osteoclastogenesis inhibitory factor.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 OS Osteoporosis.  
 FT Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /note= "Signal peptide"  
 FT Protein 22..401  
 FT /note= "Mature OCIF, Claim 6"  
 PN WO9626217-AL.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; J00374.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW ) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR N-PSDB; T36685.

PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 for bone resorption control, esp. treatment of osteoporosis  
 PS Disclosure; Page 64-66; 183pp; Japanese.  
 CC This sequence represents the full length osteoclastogenesis inhibitory  
 CC factor (OCIF) of the invention. The OCIF has a molecular weight by  
 CC SDS-PAGE of 60 kD under reducing conditions and 120 kD under non-  
 CC reducing conditions. The protein is adsorbed onto cation-exchangers  
 CC or heparin and its activity is lowered after 10 mins at 70 deg.C or  
 CC 30 mins at 56 deg.C, and is lost after 10 mins at 90 deg.C. OCIF is  
 CC useful in the control of bone resorption and therefore in the  
 CC treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 CC Sequence 401 AA;  
 Query Match 100.0%; Score 47; DB 20; Length 401;  
 Best Local Similarity 100.0%; Pred. No. 2.01e+01;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 355 yhfpk 359  
 |||||  
 QY 2 YHFPK 6

RESULT 16  
 ID R9925 standard; Protein; 401 AA.  
 AC R9925;  
 DT 22-APR-1997 (first entry)  
 DE Full length osteoclastogenesis inhibitory factor.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 OS Osteoporosis.  
 FT Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /note= "Signal peptide"  
 FT Protein 22..401  
 FT /note= "Mature OCIF, Claim 6"  
 PN WO9626217-AL.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; J00374.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW ) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR N-PSDB; T36685.

PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 for bone resorption control, esp. treatment of osteoporosis  
 PS Disclosure; Page 64-66; 183pp; Japanese.  
 CC This sequence represents the full length osteoclastogenesis inhibitory  
 CC factor (OCIF) of the invention. The OCIF has a molecular weight by  
 CC SDS-PAGE of 60 kD under reducing conditions and 120 kD under non-  
 CC reducing conditions. The protein is adsorbed onto cation-exchangers  
 CC or heparin and its activity is lowered after 10 mins at 70 deg.C or  
 CC 30 mins at 56 deg.C, and is lost after 10 mins at 90 deg.C. OCIF is  
 CC useful in the control of bone resorption and therefore in the  
 CC treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 CC Sequence 401 AA;  
 Query Match 100.0%; Score 47; DB 20; Length 401;  
 Best Local Similarity 100.0%; Pred. No. 2.01e+01;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 355 yhfpk 359  
|:|:|  
QY 2 YHFPK 6

RESULT 17  
ID P81265 standard; protein; 505 AA.  
AC P81265;  
DT 12-DEC-1990 (first entry)  
DE Sequence encoded by part of AAC(6')-APH(2'') (6'-acetyltransferase-2''-  
DE phosphotransferase) gene in the plasmid pIP800 of Streptococcus  
DE faecalis JH 102  
KW Streptococcal plasmid pGB3012; kanamycin resistant;  
KW gentamycin resistant; Staphylococcus; antibiotic resistant.  
OS Streptococcus faecalis.  
PN FR2601965-A.  
PD 29-JAN-1988.  
PF 25-JUL-1986; 010829.  
PR 25-JUL-1986; FR-010829.  
PA (CNRS) Cent Nat Rech Sci (INSP).  
PI Courvalin P;  
PI WPI: 88-079435/12.  
PI N-PSDB: n81668.  
PI New DNA fragment contg. gene for detecting resistance -  
PI to kanamycin and gentamycin from Streptococcus plasmid, and new  
PI expression prod. for making hybridisation probes  
PS Claim 6; Page 27 and Fig 3; 34pp; French.  
CS SQ in n81668 codes for a protein imparting resistance to kanamycin and  
CC gentamycin. The DNA is useful for making hybridisation probes to detect  
CC kanamycin and gentamycin resistance in clinically-isolated cell cultures.  
SQ Sequence 505 AA;

Query Match 91.5%; Score 43; DB 1; Length 505;  
Best Local Similarity 80.0%; Pred. NO. 6.35e+01;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 86 yhfpk 90  
|:|:|  
QY 2 YHFPK 6

RESULT 18  
ID R90766 standard; protein; 1137 AA.  
AC R90766;  
DT 25-APR-1996 (first entry)  
DE Tumour suppressor protein HTS-1.  
KW Tumour suppressor protein; HTS-1; cancer; therapy; prognosis.  
OS Homo sapiens.  
PN US5491064-A.  
PD 13-FEB-1996.  
PF 17-JUL-1992; 916762.  
PR 17-JUL-1992; US-916762.  
PR 05-JAN-1995; US-369043.  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
PI Howley PM, Lichy JH;  
PI WPI: 96-116319/12.  
PI N-PSDB: T12473.  
PI Human tumour suppressor gene, HTS-1 - shows differential expression  
PI in tumorigenic and normal cells and may be useful in tumour  
PI prognosis  
PS Claim 2; Column 19-26; 17pp; English.  
CC Human tumour suppressor protein HTS-1 (R90766) is produced by  
CC prokaryotic or eucaryotic host cells transformed or transfected by  
CC a vector including an isolated HTS-1 cDNA (T12473). Expression of  
CC HTS-1 is associated with tumour suppression. The purified HTS-1  
CC protein can be used to raise antibodies for use in HTS-1 detection.  
SQ Sequence 1137 AA;

Query Match 89.4%; Score 42; DB 15; Length 1137;  
Best Local Similarity 80.0%; Pred. NO. 8.43e+01;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 720 yhfpk 724  
|:|:|  
QY 2 YHFPK 6

RESULT 19  
ID R66461 standard; protein; 188 AA.  
AC R66461;  
DT 24-AUG-1995 (first entry)  
DE AF-17 protein, N-terminal region with similarity to peregrin.  
KW Acute lymphoblastic leukaemia; acute nonlymphoblastic leukaemia;  
KW chromosomal translocation; abnormality; detection;  
KW chimeric ALL-1/AF-17 protein; ALL-1 fused gene on chromosome 17.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT Region 1..118  
FT /note= "cysteine-rich region with similarity to  
FT peregrin protein"  
FT Region 45..139  
FT /note= "region of 90% identity to anonymous human  
FT cDNA sequence (Genbank Acc.No.T06113)"  
FT Domain 8..30  
FT /label= zinc-finger  
FT Domain 51..68  
FT /label= zinc-finger  
FT Domain 120..140  
FT /label= zinc-finger  
PN W09426930-A.  
PD 24-NOV-1994.  
PF 22-APR-1994; U04496.  
PR 14-MAY-1993; US-062443.  
PR (UYJE-) UNIV JEFFERSON THOMAS.  
PI Canaani E, Croce C;  
PI WPI: 95-006818/01.  
PI New acute lymphocytic leukaemia gene prods. - used for the  
PI diagnosis and treatment of leukaemias, partic. acute  
PI lymphoblastic or nonlymphoblastic leukaemia  
PS Claim 53; Page 134-135; 207pp; English.  
CC Monoclonal antibodies which bind to at least part of the chimeric  
CC ALL-1/AF-17 protein (R66459) or to part of the AF-17 protein (see  
CC R66460 and R66461) are claimed. The antibodies are useful for  
CC diagnosing acute lymphoblastic and non-lymphoblastic leukaemia.  
SQ Sequence 188 AA;

Query Match 85.1%; Score 40; DB 12; Length 188;  
Best Local Similarity 80.0%; Pred. NO. 1.47e+02;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 179 yhfsk 183  
|:|:|  
QY 2 YHFPK 6

RESULT 20  
ID R23005 standard; protein; 255 AA.  
AC R23005;  
DT 29-OCT-1992 (first entry)  
DE Protein transcribed from the mba sequence of plasmid pKDC50.  
KW Virulent; Salmonella choleraesuis; mouse bacteremia; Salmonella.  
OS Salmonella choleraesuis.  
PN J04079885-A.  
PD 13-MAR-1992.  
PF 24-JUL-1990; 194069.  
PR 24-JUL-1990; JP-194069.  
PA (KIBU-) KIBUN KK.  
PA (KITA ) KITASATO RES INST.  
PI WPI: 92-137918/17.  
PI N-PSDB: Q24046.  
PI New virulent plasmid pKDC50 - provides DNA fragment useful as  
PI probe for detecting Salmonella  
PS Disclosure; Fig 2; 9pp; Japanese.  
CC The 32 kD protein may be transcribed from the second reading frame of  
CC the "mba" fragment of virulent plasmid pKDC50 from nucleotides 2564-

CC 3331. Three other proteins may also be transcribed from the Salmonella  
 CC "mba" fragment, which is capable of causing mouse bacteremia.  
 CC See also R23004-7.  
 SQ Sequence 255 AA;

Query Match 85.1%; Score 40; DB 4; Length 255;  
 Best Local Similarity 100.0%; Pred. No. 1.47e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 135 yhhp 138  
 ||||  
 Qy 2 YHFP 5

## RESULT 21

ID R99357 standard; Protein; 390 AA.

AC R99357; 05-MAY-1997 (first entry)

Human tumour necrosis factor receptor.

Tumour necrosis factor; TNF; receptor; TNF-beta; ligand; tumour;  
 KW differentiation; immune response; autoimmune disease; inflammation;  
 KW septic shock; graft-versus-host; apoptosis.

OS Homo sapiens.

PH Key Location/Qualifiers

FT Peptide 1..21

FT /label= sig\_peptide

PN WO9628546-A1.

PD 19-SEP-1996.

PF 15-MAR-1995; U03216.

PR 15-MAR-1995; WO-U03216.

PR 29-MAR-1995; ZA-002587.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Fleischmann RD, Greene JM;

DR WPI; 96-433821/43.

DR N-PSDB; T35475.

PT New human tumour necrosis factor receptor - used to develop prods.

PT for treating e.g. tumours, infection, auto:immune disease, graft

PT rejection, cytotoxicity or inflammation

PS Claim 14; Fig 1; 59pp; English.

CC The receptor binds to TNF, and in particular, TNF-beta.

CC The receptor may be used for screening for antagonists and agonists

CC of the receptor and for ligands for the receptor. Such agonists may

CC be used to inhibit the growth of tumours, to stimulate cellular

CC differentiation, to mediate the immune response and anti-viral

CC response, to regulate growth and provide resistance to certain

CC infections. The antagonists may be used therapeutically, to treat

CC autoimmune diseases, inflammation, septic shock, to inhibit graft-

CC versus-host reactions, and to prevent apoptosis.

SQ Sequence 390 AA;

Query Match 85.1%; Score 40; DB 20; Length 390;

Best Local Similarity 100.0%; Pred. No. 1.47e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 355 yhhp 358

||||

Qy 2 YHFP 5

## RESULT 22

ID W00729 standard; Protein; 422 AA.

AC W00729;

DT 03-JAN-1997 (first entry)

DE Daunorubicin 14-hydroxylase.

KW dxra; doxorubicin; bioconversion; anthracycline; antitumour therapy;

KW daunorubicin 14-hydroxylase.

OS Streptomyces peucetius strain 29050.

PN WO9627014-A1.

PD 06-SEP-1996.

PF 20-FEB-1996; E00692.

PR 27-FEB-1995; US-396218.

PA (PHAA) PHARMACIA SPA.

PI Brene U, Colombo AL, Hutchinson CR, Otten S, Scotti C;

PI Solari INVENTI A, Inventa AS;  
 DR WPI; 96-412780/41.

DR N-PSDB; T33617.

PT New DNA encoding daunorubicin 14-hydroxylase - useful for converting

PT daunorubicin to the antitumour agent doxorubicin

PS Claim 4; Page 15-18; 31pp; English.

CC The present sequence is that of the daunorubicin 14-hydroxylase, isolated

CC from Streptomyces peucetius strain 29050. The enzyme converts

CC daunorubicin to doxorubicin, an anthracycline used widely as an

CC antitumour agent. A vector contg. the dxra gene (T33617) could be used to

CC transform a suitable host cell which produces daunorubicin, in order to

CC enhance bioconversion of daunorubicin to doxorubicin. Bioconversion can

CC be carried out either by isolating the free or immobilised

CC transformed cells or by isolating the enzyme, which can be used in the

CC free form or immobilised to resins, glass, cellulose or similar

CC substances by ionic or covalent bonds, or grafted to fibres permeable to

CC the substrate or insolubilised by cross-linkage.

SQ Sequence 422 AA;

Query Match 85.1%; Score 40; DB 19; Length 422;

Best Local Similarity 100.0%; Pred. No. 1.47e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 156 yhhp 159

||||

Qy 2 YHFP 5

## RESULT 23

ID R91277 standard; Protein; 527 AA.

AC R91277; 03-OCT-1996 (first entry)

DE Exophiala spinifera fumonisin esterase.

KW Fumonisin; esterase; transgenic plant; recombinant microorganism;

KW expression; probiotic; feed inoculant; degradation; detoxification;

KW maize seed; grain; animal feed.

OS Exophiala spinifera.

PN WO9606175-A2.

PD 29-FEB-1996.

PF 11-AUG-1995; U10284.

PR 12-AUG-1994; US-289595.

PR 07-JUN-1995; US-484815.

PA (PION-) PIONEER HI-BRED INT INC.

PI Durick J, Rood TA;

DR WPI; 96-151378/15.

PT Detoxification of fumonisins and related mycotoxin cpds. in grains -

PT using an enzyme esp. an esterase, from Exophiala spinifera,

PT Rhinocladella atrovirens or a bacterium

PS Example 8; Pages 33-34; 54pp; English.

CC The present sequence is the Exophiala spinifera (ATCC 74269)

CC fumonisin esterase, which was isolated from a maize seed. The gene

CC which encodes the esterase can be used to produce transgenic

CC plants and genetically engineered microorganisms, capable of

CC expressing the esterase. The microorganisms can be used as a

CC probiotic or feed inoculant, along with the esterase to degrade

CC fumonisins and related cpds. partic. for the detoxification of

CC maize seed pre- or post-harvest (i.e. during the storage or

CC processing of the harvested grain, or in the processed grain)

CC prior to its use as an animal feed.

SQ Sequence 527 AA;

Query Match 85.1%; Score 40; DB 18; Length 527;

Best Local Similarity 80.0%; Pred. No. 1.47e+02;

Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 256 ynhfk 260

||||

Qy 2 YHFPK 6

## RESULT 24

ID R14402 standard; Protein; 934 AA.

AC R14402;



DT 14-FEB-1992 (first entry)  
 DE Soluble Insulin-like Growth Factor receptor.  
 KW Extracellular receptor; IGF.  
 PN WO9117253-A.  
 PD 14-NOV-1991.  
 PF 30-APR-1991; DK01116.  
 PR 30-APR-1990; DK-001064.  
 PA (NOVO ) NOVO NORDISK A/S.  
 PI Wiberg FC, Olsen F, Boel E, Kjeldsen TB, Moller NPH.  
 DR WPI; 91-353774/48.  
 DR N-PSDB:Q14380.  
 PT Mammalian cells contg. hybrid DNA for extracellular receptor -  
 PT including exons from two separate receptors, having improved  
 PT growth performance in serum-free medium  
 PS Disclosure; Fig 3; 57pp; English.  
 CC The amino acid sequence is that of the soluble insulin-like growth  
 CC factor (IGF) receptor. It is used here as part of a fusion protein  
 CC produced by a hybrid DNA insert comprising a DNA sequence encoding part  
 CC of the extracellular domain of the IR and a sequence coding for part of  
 CC the ECD of an insulin like growth factor (IGF). Cells contg. this insert  
 CC have better growth characteristics in both serum-contg. and partic.  
 CC serum-free media. This makes them useful for expressing polypeptides  
 CC (e.g. human blood clotting factors; tissue plasminogen activator;  
 CC human protein C), encoded by a third DNA sequence. The polypeptides  
 CC can now be purified more easily and produced in better yield. See  
 CC also R14403.  
 SQ Sequence 934 AA;

Query Match 85.1%; Score 40; DB 3; Length 934;  
 Best Local Similarity 80.0%; Pred. No. 1.47e+02;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 75 yrpk 79  
 I:||||  
 QY 2 YHFPK 6

RESULT 25  
 ID R15048 standard; Protein: 934 AA.  
 AC R15048;  
 DT 14-FEB-1992 (first entry)  
 DE Soluble human IGF-I receptor.  
 KW Insulin-like growth factor.  
 OS Homo sapiens.  
 PN WO9117252-A.  
 PD 14-NOV-1991.  
 PF 30-APR-1991; DK01115.  
 PR 30-APR-1990; DK-001063.  
 PR 03-AUG-1990; DK-001855.  
 PA (NOVO ) NOVO NORDISK A/S.  
 PI Kjeldsen T, Andersen AS, Wiberg F, Moller N, Rasmussen J;  
 WPI; 91-353773/48.  
 DR N-PSDB: Q14814.  
 PT New hybrid cellular receptor - has 2 sequences encoding  
 PT extracellular domain of receptors having different ligand binding  
 PT specificity e.g. insulin- and IGF-receptors.  
 PS Example 1; Fig 4; 97pp; English.  
 CC The sequence was deduced from a cDNA sequence prep. from mRNA  
 CC isolated from human term placenta by PCR amplification. The full  
 CC sequence was assembled from four subcloned fragments and was found  
 CC to be identical to that published by Ullrich et al (The EMBO  
 CC Journal 5(120), 1986, pp 2503-2512). The DNA can be ligated to DNA  
 CC encoding a sol. insulin receptor (see Q14813) to prepare hybrid  
 CC proteins. The new receptors are useful for screening for ligands  
 CC and for localising the ligand binding site. They are also useful  
 CC for establishing the 3-D structure of the ligand binding site and  
 CC for designing analogues or functional equivalents.  
 CC See also R15047-51.  
 SQ Sequence 934 AA;

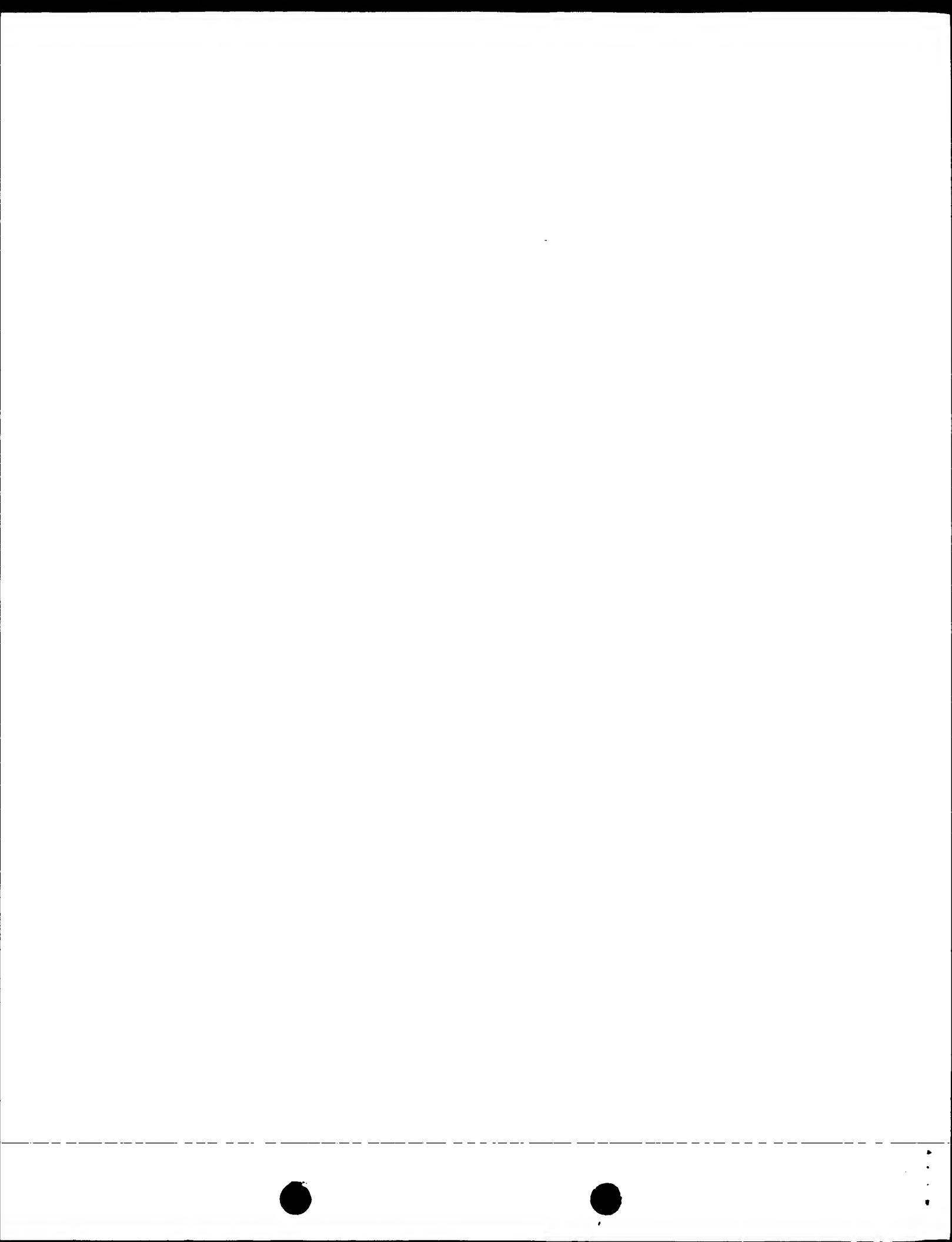
Query Match 85.1%; Score 40; DB 3; Length 934;  
 Best Local Similarity 80.0%; Pred. No. 1.47e+02;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 75 yrpk 79  
 I:||||  
 QY 2 YHFPK 6  
 RESULT 26  
 ID R14403 standard; Protein: 948 AA.  
 AC R14403;  
 DT 14-FEB-1992 (first entry)  
 DE Extracellular domain of hybrid insulin/IGF-I receptor.  
 KW Extracellular receptor; IGF; fusion protein.  
 PN WO9117253-A.  
 PD 14-NOV-1991.  
 PF 30-APR-1991; DK01116.  
 PR 30-APR-1990; DK-001064.  
 PA (NOVO ) NOVO NORDISK A/S.  
 PI Wiberg FC, Olsen F, Boel E, Kjeldsen TB, Moller NPH.  
 DR WPI; 91-353774/48.  
 DR N-PSDB:Q14380.  
 PT Mammalian cells contg. hybrid DNA for extracellular receptor -  
 PT including exons from two separate receptors, having improved  
 PT growth performance in serum-free medium  
 PS Disclosure; Fig 4; 57pp; English.  
 CC The amino acid sequence is that encoded by a hybrid DNA insert comprising  
 CC a DNA sequence encoding part of the extracellular domain of the IR and  
 CC a sequence coding for part of the ECD of an insulin like growth factor  
 CC (IGF). Cells contg. this insert have better growth characteristics in  
 CC both serum-contg. and partic. serum-free media. This makes them useful  
 CC for expressing polypeptides (e.g. human blood clotting factors; tissue  
 CC plasminogen activator; human protein C), encoded by a third DNA  
 CC sequence. The polypeptides can now be purified more easily and  
 CC produced in better yield. See also R14402.  
 SQ Sequence 948 AA;

Query Match 85.1%; Score 40; DB 3; Length 948;  
 Best Local Similarity 80.0%; Pred. No. 1.47e+02;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 73 yrpk 77  
 I:||||  
 QY 2 YHFPK 6

Search completed: Wed Dec 3 10:06:32 1997  
 Job time : 12 secs.



\*\*\*\*\*  
W P E R E H  
\*\*\*\*\*  
(TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Tue Dec 2 14:43:12 1997; MasPar time 5.65 Seconds  
30.699 Million cell updates/sec

ular output not generated.

Title: >US-08-915-004-1  
Description: (1-6) from US08915004.pep  
Perfect Score: 47  
Sequence: 1 XYHFK 6

Scoring table: GAP 150  
PAM 15  
Searched: 91006 seqs, 28888923 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: p1r51  
1:ann1 2:ann2 3:ann3 4:ann4 5:unann1 6:unann2 7:unann3  
8:unann4 9:unann5 10:unann6 11:unann7 12:unann8  
13:unann9 14:unann10 15:unenc 16:unrev  
Statistics: Mean 20.106; Variance 27.499; scale 0.731

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

sult No.	Score	Query Match	Length	ID	Description	Pred. No.
1	47	100.0	103	14	collagen alpha 1(XVI)	4.96e+00
2	47	100.0	562	14	collagen alpha 1(XVI)	4.96e+00
3	43	91.5	479	9	aminoglycoside acety	2.88e+01
4	43	91.5	479	9	aminoglycoside resis	2.88e+01
5	43	91.5	803	14	diptidyl aminopept	2.88e+01
6	43	91.5	803	13	diptidyl aminopept	2.88e+01
7	43	91.5	865	13	diptidyl aminopept	2.88e+01
8	43	91.5	1014	11	HEX2 protein - yeast	2.88e+01
9	42	89.4	109	16	nonstructural protei	4.41e+01
10	42	89.4	109	16	nonstructural protei	4.41e+01
11	42	89.4	109	16	nonstructural protei	4.41e+01
12	42	89.4	109	4	nonstructural protei	4.41e+01
13	42	89.4	228	12	PES-1B protein - Ca	4.41e+01
14	42	89.4	264	12	PES-1 protein - Ca	4.41e+01
15	42	89.4	336	9	replication protein	4.41e+01
16	42	89.4	349	16	hypothetical protein	4.41e+01
17	42	89.4	349	11	hypothetical protein	4.41e+01
18	42	89.4	645	13	HTS1 - human (fragm	4.41e+01
19	42	89.4	700	12	gene unc-93 protein	4.41e+01
20	42	89.4	700	12	gene unc-93 protein	4.41e+01
21	42	89.4	1511	12	probable membrane pr	4.41e+01

22	41	87.2	238	9	I39731	hydrogenase small ch	6.70e+01
23	41	87.2	252	10	B32053	hypothetical protein	6.70e+01
24	40	85.1	113	16	S22627	hypothetical protein	1.01e+02
25	40	85.1	209	16	S62589	hypothetical protein	1.01e+02
26	40	85.1	271	4	IRECM	colicin M - Escheric	1.01e+02
27	40	85.1	283	11	S46770	hypothetical protein	1.01e+02
28	40	85.1	327	10	B42653	36K protein - Achole	1.01e+02
29	40	85.1	348	4	MFNZS	matrix protein - Sen	1.01e+02
30	40	85.1	348	4	MFNZSV	matrix protein - Sen	1.01e+02
31	40	85.1	351	8	B49906	rFB homolog - Xanth	1.01e+02
32	40	85.1	353	9	E64023	hypothetical protein	1.01e+02
33	40	85.1	361	9	I35054	rDTP-D-glucose 4,6-d	1.01e+02
34	40	85.1	361	9	B5213	rFB protein - Shige	1.01e+02
35	40	85.1	361	9	S15299	rFB protein - Salmo	1.01e+02
36	40	85.1	449	8	A45709	major early-transcri	1.01e+02
37	40	85.1	484	6	S00757	deoxyribodipyrimidin	1.01e+02
38	40	85.1	531	11	S41986	nuclear protein SON1	1.01e+02
39	40	85.1	624	10	A52220	penicillin-binding p	1.01e+02
40	40	85.1	692	1	PABY12	phosphoprotein phosph	1.01e+02
41	40	85.1	1020	14	S05944	neuronal cell surfac	1.01e+02
42	40	85.1	1021	14	A57112	contactin precursor	1.01e+02
43	40	85.1	1289	1	RMXRR3	mRNA guanylyltransfe	1.01e+02
44	40	85.1	1367	1	IGHUR1	insulin-like growth	1.01e+02
45	40	85.1	1371	5	A33837	insulin-like growth	1.01e+02

ALIGNMENTS

RESULT 1  
ENTRY A56101 #type fragment  
TITLE collagen alpha 1(XVIII) chain precursor short form - mouse  
ORGANISM #formal\_name Mus musculus #common\_name house mouse  
DATE 03-Oct-1995 #sequence\_revision 03-Oct-1995 #text\_change 03-Oct-1995  
ACCESSIONS A56101  
REFERENCE A56101  
#authors Rehn, M.; Pihlajaniemi, T.  
#journal J. Biol. Chem. (1995) 270:4705-4711  
#title Identification of three N-terminal ends of type XVIII collagen chains and tissue-specific differences in the expression of the corresponding transcripts. The longest form contains a novel motif homologous to rat and Drosophila frizzled proteins.

#accession A56101  
#status preliminary  
#molecule\_type mRNA  
#residues 1-103 #label REH  
##cross-references GB:U11636

GENETICS COL18A1  
SUMMARY #length 103 #checksum 5841

Query Match 100.0%; Score 47; DB 14; Length 103;  
Best Local Similarity 100.0%; Pred. No. 4.96e+00;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 74 yhfpk 78  
Qy 2 YHFK 6

RESULT 2  
ENTRY B56101 #type fragment  
TITLE collagen alpha 1(XVIII) chain precursor long form - mouse  
ORGANISM #formal\_name Mus musculus #common\_name house mouse  
DATE 03-Oct-1995 #sequence\_revision 03-Oct-1995 #text\_change 03-Oct-1995  
ACCESSIONS B56101  
REFERENCE A56101  
#authors Rehn, M.; Pihlajaniemi, T.  
#journal J. Biol. Chem. (1995) 270:4705-4711

#title Identification of three N-terminal ends of type XVIII collagen chains and tissue-specific differences in the expression of the corresponding transcripts. The longest form contains a novel motif homologous to rat and Drosophila frizzled proteins.

#accession B56101  
#status preliminary  
#molecule\_type mRNA  
#residues 1-562 ##label REH  
##cross-references GB:U11637

## GENETICS

#gene  
#summary

COL18A1  
#length 562 #checksum 115

Query Match 100.0%; Score 47; DB 14; Length 562;  
Best Local Similarity 100.0%; Pred. No. 4.96e+00;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

533 yhpfpk 537  
|||||  
QY 2 YHFPK 6

## RESULT

ENTRY #type complete  
TITLE aminoglycoside acetyltransferase - Enterococcus faecalis  
ORGANISM #formal\_name Enterococcus faecalis  
DATE 05-Oct-1988 #sequence\_revision 05-Oct-1988 #text\_change 18-Jun-1993

## ACCESSIONS

## REFERENCE

#authors Ferretti, J.J.; Gilmore, K.S.; Courvalin, P.  
#journal J. Bacteriol. (1986) 167:631-638  
#title Nucleotide sequence analysis of the gene specifying the bifunctional 6'-aminoglycoside acetyltransferase 2'-aminoglycoside phosphotransferase enzyme in Streptococcus faecalis and identification and cloning of gene regions specifying the two activities.  
#cross-references MUID:86277923

#accession A26048  
#molecule\_type DNA  
#residues 1-479 ##label FER

#note the authors translated the codon TAT for residue 65 as Thr and GGT for residue 366 as Glu  
antibiotic resistance  
#length 479 #molecular-weight 56855 #checksum 5383

## KEYWORDS

## SUMMARY

Query Match 91.5%; Score 43; DB 9; Length 479;  
Best Local Similarity 80.0%; Pred. No. 2.88e+01;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 86 yhpfpk 90  
|||||  
QY 2 YHFPK 6

## RESULT

ENTRY #type complete  
TITLE aminoglycoside resistance protein aacA-aphD - Staphylococcus aureus transposon Tn4001  
ORGANISM #formal\_name Staphylococcus aureus  
DATE 03-May-1994 #sequence\_revision 20-Feb-1995 #text\_change 20-Feb-1995

## ACCESSIONS

## REFERENCE

#authors Rouch, D.A.; Byrne, M.E.; Kong, Y.C.; Skurray, R.A.  
#journal J. Gen. Microbiol. (1987) 133:3039-3052  
#title The aacA-aphD gentamicin and kanamycin resistance determinant of Tn4001 from Staphylococcus aureus: expression and nucleotide sequence analysis.  
#accession S26353  
#molecule\_type DNA  
#residues 1-479 ##label ROU

##cross-references EMBL:M18086

##note the authors translated the codon GAT for residue 374 as Asp

KEYWORDS antibiotic resistance

SUMMARY #length 479 #molecular-weight 56855 #checksum 5383

Query Match 91.5%; Score 43; DB 9; Length 479;  
Best Local Similarity 80.0%; Pred. No. 2.88e+01;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 86 yhpfpk 90  
|||||  
QY 2 YHFPK 6

## RESULT

ENTRY #type complete  
TITLE dipeptidyl aminopeptidase-like protein (EC 3.4.14.-) short form, DPPX-S - bovine  
ORGANISM #formal\_name Bos primigenius taurus #common\_name cattle  
DATE 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 15-Nov-1996

## ACCESSIONS

## REFERENCE

#authors Wada, K.; Yokotani, N.; Hunter, C.; Doi, K.; Wenthold, R.J.; Shimasaki, S.  
#journal Proc. Natl. Acad. Sci. U.S.A. (1992) 89:197-201  
#title Differential expression of two distinct forms of mRNA encoding members of a dipeptidyl aminopeptidase family.  
#cross-references MUID:92108018

## #accession

A41793

##status preliminary; not compared with conceptual translation  
##molecule\_type mRNA  
##residues 1-803 ##label WAD  
##cross-references NCBI:75138  
##note sequence extracted from NCBI backbone  
dipeptidylpeptide hydrolase

## KEYWORDS

## FEATURE

257,342  
#binding\_site carbohydrate (Asn) (covalent) #status predicted

SUMMARY #length 803 #molecular-weight 90459 #checksum 9315

Query Match 91.5%; Score 43; DB 14; Length 803;  
Best Local Similarity 80.0%; Pred. No. 2.88e+01;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 280 yhpfpk 284  
|||||  
QY 2 YHFPK 6

## RESULT

ENTRY #type complete  
TITLE dipeptidyl aminopeptidase like protein - human  
ORGANISM #formal\_name Homo sapiens #common\_name man  
DATE 06-Sep-1996 #sequence\_revision 06-Sep-1996 #text\_change 06-Sep-1996

## ACCESSIONS

## REFERENCE

I68600  
Yokotani, N.; Doi, K.; Wenthold, R.J.; Wada, K.  
Hum. Mol. Genet. (1993) 2:1037-1039  
#title Non-conservation of a catalytic residue in a dipeptidyl aminopeptidase IV-related protein encoded by a gene on human chromosome 7.  
#cross-references MUID:93372805

## #accession

I68600

##status preliminary; translated from GB/EMBL/DBJ  
##molecule\_type mRNA  
##residues 1-803 ##label RES  
##cross-references GB:M96860; MUID:9306707; CDS\_PID:9306708  
#length 803 #molecular-weight 91355 #checksum 814

## SUMMARY

Query Match 91.5%; Score 43; DB 13; Length 803;

```

Best Local Similarity 80.0%; Pred. No. 2.88e+01;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 280 yhyphk 284
QY 2 YHFPK 6

RESULT 7
ENTRY dipeptidyl aminopeptidase like protein - human
TITLE dipeptidyl aminopeptidase like protein - human
ORGANISM Homo sapiens
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
ACCESSIONS I54331
REFERENCE I54331
#authors Yokotani, N.; Doi, K.; Wenthold, R.J.; Wada, K.
#journal Hum. Mol. Genet. (1993) 2:1037-1039
#title Non-conservation of a catalytic residue in a dipeptidyl
aminopeptidase IV-related protein encoded by a gene on
human chromosome 7.
#cross-references MUID:93372805
#accession I54331
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-865 #label RES
#cross-references GB:M96859; NID:g306705; CDS.PID:g306706
SUMMARY #length 865 #molecular-weight 97587 #checksum 6089

Query Match 91.5%; Score 43; DB 13; Length 865;
Best Local Similarity 80.0%; Pred. No. 2.88e+01;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 342 yhyphk 346
QY 2 YHFPK 6

RESULT 8
ENTRY HEX2 protein - yeast (Saccharomyces cerevisiae)
TITLE protein D3402; protein P2F240; protein Y09813.06c; protein
ALTERNATE_NAMES IDRO28c; SRN1 protein
ORGANISM Saccharomyces cerevisiae
DATE 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
ACCESSIONS S32613; S50935; S31242; S17254; S27374; S63438; S67842;
S67841; A44440; B44440
REFERENCE S32613
#authors Tung, K.S.; Norbeck, L.L.; Nolan, S.L.; Atkinson, N.S.;
Hopper, A.K.
#submission submitted to the EMBL Data Library, December 1992
#accession S32613
#molecule_type DNA
#residues 1-1014 #label T01
#cross-references EMBL:M90540
REFERENCE S50930
#authors Bowman, S.
#submission submitted to the EMBL Data Library, January 1994
#accession S50935
#molecule_type DNA
#residues 1-1014 #label BOW
#cross-references EMBL:247814
REFERENCE A44440
#authors Tung, K.S.; Norbeck, L.L.; Nolan, S.L.; Atkinson, N.S.;
Hopper, A.K.
#journal Mol. Cell. Biol. (1992) 12:2673-2680
#title SRN1, a yeast gene involved in RNA processing, is identical
to HEX2/REG1, a negative regulator in glucose repression.
#cross-references MUID:92269840
#accession S31242
#molecule_type DNA
#residues 962-1014 #label T02

```

---

```

#cross-references EMBL:M90540
REFERENCE S17254
#authors Niederacher, D.; Entian, K.D.
#journal Eur. J. Biochem. (1991) 200:311-319
#title Characterization of Hex2 protein, a negative regulatory
element necessary for glucose repression in yeast.
#cross-references MUID:91364678
#accession S17254
#molecule_type DNA
#residues 1-375, 'EE', 378-533, 'K', 535-656, 'H', 658-987,
'QEVQWASTCTLGKRVTSKKKMENTAVRRKNFEVN' #label NIE
#cross-references EMBL:M33703
REFERENCE S27374
#authors Niederacher, D.; Entian, K.D.
#submission submitted to the EMBL Data Library, April 1990
#accession S27374
#molecule_type DNA
#residues 1-375, 'EE', 378-533, 'K', 535-656, 'H', 658-988, 'T', 890-987,
'QEVQWASTCTLGKRVTSKKKMENTAVRRKNFEVN' #label NI2
#cross-references EMBL:M33703
REFERENCE S63416
#authors Eide, L.G.; Sander, C.; Prydz, H.
#submission submitted to the EMBL Data Library, February 1996
#description Sequencing and analysis of a 35.4 kb region on the left arm
of chromosome IV for Saccharomyces cerevisiae reveal 23
open reading frames.
#accession S63438
#molecule_type DNA
#residues 775-1014 #label EID
#cross-references EMBL:X95966
REFERENCE S67842
#authors Arnold, W.; Becker, A.; Jaeger, W.; Kuester, H.; Nussbaumer,
B.
#submission submitted to the Protein Sequence Database, July 1996
#accession S67842
#molecule_type DNA
#residues 1-1014 #label ARN
#cross-references EMBL:274324
#experimental_source strain S288C
REFERENCE S67822
#authors Prydz, H.; Eide, L.G.
#submission submitted to the Protein Sequence Database, July 1996
#accession S67841
#molecule_type DNA
#residues 775-1014 #label PRY
#cross-references EMBL:274324
#experimental_source strain S288C
GENETICS
#gene LISTA:HEX2; SRN1
#map_position 4R
#KEYWORDS nucleus; regulatory protein
SUMMARY #length 1014 #molecular-weight 112615 #checksum 705

Query Match 91.5%; Score 43; DB 11; Length 1014;
Best Local Similarity 80.0%; Pred. No. 2.88e+01;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 338 yhyphk 342
QY 2 YHFPK 6

RESULT 9
ENTRY nonstructural protein, 12.8K - porcine hemagglutinating
TITLE encephalomyelitis virus (strain 67N)
ORGANISM porcine encephalomyelitis virus
DATE 13-Jan-1996 #sequence_revision 13-Jan-1996 #text_change
ACCESSIONS S58182
REFERENCE S58182
#authors Vieler, E.; Schlapp, T.; Herbst, W.
#submission submitted to the EMBL Data Library, July 1995

```

```
#accession S58182
##status preliminary
##residues 1-109 ##label VIE
##cross-references EMBL:X89861
SUMMARY #length 109 #molecular-weight 12818 #checksum 4120

Query Match 89.4%; Score 42; DB 16; Length 109;
Best Local Similarity 80.0%; Pred. No. 4.41e+01;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 31 yqfpk 35
Qy 2 YHFPK 6

RESULT 10
ENTRY nonstructural protein, 12.8K - porcine hemagglutinating
TITLE encephalomyelitis virus (strain VM572)
ORGANISM #formal_name porcine hemagglutinating encephalomyelitis virus
DATE 13-Jan-1996 #sequence_revision 13-Jan-1996 #text_change
13-Jan-1996

ACCESSIONS S58186
REFERENCE Vieler, E.; Schlapp, T.; Herbst, W.
#authors submitted to the EMBL Data Library, July 1995
#submission S58186
#accession
##status preliminary
##residues 1-109 ##label VIE
##cross-references EMBL:X89862
SUMMARY #length 109 #molecular-weight 12773 #checksum 4821

Query Match 89.4%; Score 42; DB 16; Length 109;
Best Local Similarity 80.0%; Pred. No. 4.41e+01;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 31 yqfpk 35
Qy 2 YHFPK 6

RESULT 11
ENTRY nonstructural protein, 12.8K - porcine hemagglutinating
TITLE encephalomyelitis virus (strain N19)
ORGANISM #formal_name porcine hemagglutinating encephalomyelitis virus
DATE 13-Jan-1996 #sequence_revision 13-Jan-1996 #text_change
13-Jan-1996

ACCESSIONS S58184
REFERENCE Vieler, E.; Schlapp, T.; Herbst, W.
#authors submitted to the EMBL Data Library, July 1995
#submission S58184
#accession
##status preliminary
##residues 1-109 ##label VIE
##cross-references EMBL:X89863
SUMMARY #length 109 #molecular-weight 12818 #checksum 4120

Query Match 89.4%; Score 42; DB 16; Length 109;
Best Local Similarity 80.0%; Pred. No. 4.41e+01;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 31 yqfpk 35
Qy 2 YHFPK 6

RESULT 12
ENTRY nonstructural protein NS2 - human coronavirus (strain OC43)
TITLE #formal_name human coronavirus
ORGANISM #formal_name human coronavirus (strain OC43)
DATE 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change
17-Feb-1994
```

---

```
ACCESSIONS A44275
REFERENCE A44275
#authors Mounir, S.; Talbot, P.J.
#journal Virology (1993) 192:355-360
#title Human coronavirus OC43 RNA 4 lacks two open reading frames
#accession A44275
##molecule_type mRNA
##residues 1-109 ##label MOU
##cross-references GB:M9576
CLASSIFICATION #superfamily bovine coronavirus nonstructural protein NS2
KEYWORDS nonstructural protein
SUMMARY #length 109 #molecular-weight 12935 #checksum 1613

Query Match 89.4%; Score 42; DB 4; Length 109;
Best Local Similarity 80.0%; Pred. No. 4.41e+01;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 31 yqfpk 35
Qy 2 YHFPK 6

RESULT 13
ENTRY PES-1B protein - Caenorhabditis elegans
TITLE #formal_name Caenorhabditis elegans
ORGANISM #formal_name Caenorhabditis elegans
DATE 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change
07-Jul-1995

ACCESSIONS S42791
REFERENCE Hope, I.A.
#authors submitted to the EMBL Data Library, November 1993
#submission S42791
#accession
##status preliminary
##molecule_type mRNA
##residues 1-228 ##label HOP
##cross-references EMBL:Z28376
CLASSIFICATION #superfamily fork head DNA-binding domain homology
FEATURE 57-147
SUMMARY #length 228 #molecular-weight 25752 #checksum 8289

Query Match 89.4%; Score 42; DB 12; Length 228;
Best Local Similarity 80.0%; Pred. No. 4.41e+01;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 222 fhfpk 226
Qy 2 YHFPK 6

RESULT 14
ENTRY PES-1 protein - Caenorhabditis elegans
TITLE PES-1A protein
ALTERNATE_NAMES #formal_name Caenorhabditis elegans
ORGANISM #formal_name Caenorhabditis elegans
DATE 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change
01-Dec-1995

ACCESSIONS S42793; S42792
REFERENCE S42791
#authors Hope, I.A.
#submission submitted to the EMBL Data Library, November 1993
#accession S42793
##molecule_type DNA
##residues 1-264 ##label HOP
##cross-references EMBL:Z28375
#accession S42792
##molecule_type mRNA
##residues 1-264 ##label HO2
##cross-references EMBL:Z28377
GENETICS
#introns 33/1; 62/1; 83/1; 162/1
```

CLASSIFICATION #superfamily fork head DNA-binding domain homology  
 FEATURE  
 93-183 #domain fork head DNA-binding domain homology #label FHD  
 SUMMARY #length 264 #molecular-weight 29480 #checksum 3098

Query Match 89.4%; Score 42; DB 12; Length 264;  
 Best Local Similarity 80.0%; Pred. No. 4.41e+01;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 258 fhfpg 262  
 :||||  
 QY 2 YHFPK 6

RESULT 15  
 ENTRY A47092 #type complete  
 TITLE replication protein repA - Enterococcus faecalis plasmid pAD1  
 ORGANISM #formal\_name Enterococcus faecalis  
 DATE 26-May-1994 #sequence\_revision 26-May-1994 #text\_change 26-May-1994

ACCESSIONS  
 A47092  
 REFERENCE  
 A47092  
 #authors Weaver, K.E.; Clewell, D.B.; An, F.  
 #journal J. Bacteriol. (1993) 175:1900-1909  
 #title Identification, characterization, and nucleotide sequence of a region of Enterococcus faecalis pheromone-responsive plasmid pAD1 capable of autonomous replication.

#accession A47092  
 #status preliminary  
 #molecule\_type DNA  
 #residues 1-336 #label WEA  
 #cross-references GB:L01794

GENETICS  
 #gene repA  
 #genome plasmid  
 SUMMARY #length 336 #molecular-weight 39031 #checksum 4594

Query Match 89.4%; Score 42; DB 9; Length 336;  
 Best Local Similarity 80.0%; Pred. No. 4.41e+01;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 18 ycfpk 22  
 :||||  
 QY 2 YHFPK 6

Search completed: Tue Dec 2 14:43:23 1997  
 Job time : 11 secs.





\*\*\*\*\*

(TM)

\*\*\*\*\*

Release 2.1d John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
 Distribution rights by Intelligenetics, Inc.

MPerch\_pp protein - protein database search, using Smith-Waterman algorithm  
 Run on: Tue Dec 2 14:42:45 1997; Maspar time 3.82 Seconds  
 33.352 Million cell updates/sec  
 Molecular output not generated.

File: >US-08-915-004-1  
 Description: (1-6) from US08915004.pep  
 Perfect Score: 47  
 Sequence: 1 XYHFPK 6

Scoring table: PAM 150  
Gap 15

Searched: 59021 seqs, 21210388 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: swiss-prot34  
 1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
 8:part8 9:part9 10:part10 11:part11

Statistics: Mean 20.932; Variance 22.971; scale 0.911

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	47	100.0	1315	2	COLLAGEN ALPHA 1(XVII)	9.76e+01
2	43	91.5	479	1	6'-AMINOGLYCOSIDE N-A	7.96e+00
3	43	91.5	859	3	DIPEPTIDYL PEPTIDASE	7.96e+00
4	43	91.5	863	3	DIPEPTIDYL PEPTIDASE	7.96e+00
5	43	91.5	865	3	DIPEPTIDYL PEPTIDASE	7.96e+00
6	43	91.5	1014	4	HEX2 YEAST	7.96e+00
7	42	89.4	109	10	NONSTRUCTURAL PROTEIN PR	1.32e+01
8	42	89.4	349	11	YJY9 YEAST	1.32e+01
9	41	87.2	252	11	YCR2-BACTK	2.16e+01
10	40	85.1	101	9	TRAL-SALTY	3.51e+01
11	40	85.1	209	11	YAL4-SCHPO	3.51e+01
12	40	85.1	255	10	VRP1-SALEN	3.51e+01
13	40	85.1	255	10	VRP1-SALDU	3.51e+01
14	40	85.1	255	10	VRP1-SALCH	3.51e+01
15	40	85.1	262	2	CON3-ANOAG	3.51e+01
16	40	85.1	283	11	YHWO-YEAST	3.51e+01
17	40	85.1	291	11	YLF4-CABEL	3.51e+01
18	40	85.1	323	9	SYK3-HAEIN	3.51e+01
19	40	85.1	325	9	SYK3-ECOLI	3.51e+01
20	40	85.1	327	7	ODEB-ACHLA	3.51e+01
21	40	85.1	337	11	YJFQ-ECOLI	3.51e+01
22	40	85.1	348	10	VMAT_SENDS	3.51e+01

23	40	85.1	348	10	VMAT_SENDF	MATRIX PROTEIN.	3.51e+01
24	40	85.1	348	10	VMAT_SENDH	MATRIX PROTEIN.	3.51e+01
25	40	85.1	348	10	VMAT_SENDZ	MATRIX PROTEIN.	3.51e+01
26	40	85.1	353	11	YC54-HAEIN	HYPOTHETICAL PROTEIN	3.51e+01
27	40	85.1	356	11	YD6B-SCHPO	HYPOTHETICAL 42.2 KD	3.51e+01
28	40	85.1	356	11	YJFR-ECOLI	HYPOTHETICAL 40.3 KD	3.51e+01
29	40	85.1	383	11	YJHO-ECOLI	HYPOTHETICAL 41.7 KD	3.51e+01
30	40	85.1	387	1	AMPL-YEAST	METHIONINE AMINOPEPTI	3.51e+01
31	40	85.1	393	11	YN8S-YEAST	HYPOTHETICAL 44.5 KD	3.51e+01
32	40	85.1	394	1	AMPL-HUMAN	POTATIVE METHIONINE A	3.51e+01
33	40	85.1	439	2	CLUS-BOVIN	CLUSTERIN PRECURSOR	3.51e+01
34	40	85.1	449	6	ME53-NPVAC	EARLY 53 KD PROTEIN	3.51e+01
35	40	85.1	483	7	PHR-ANANI	DEOXYRIBODIPYRIMIDINE	3.51e+01
36	40	85.1	492	9	SYK_THETH	LYSYL-TRNA SYNTHETASE	3.51e+01
37	40	85.1	523	5	KPT2-HUMAN	SERINE/THREONINE-PROT	3.51e+01
38	40	85.1	531	9	SON1-YEAST	NUCLEAR PROTEIN SON1	3.51e+01
39	40	85.1	555	11	YNV9-YEAST	HYPOTHETICAL 63.8 KD	3.51e+01
40	40	85.1	682	3	E13B-BACCI	GLUCAN ENDO-1,3-BETA-	3.51e+01
41	40	85.1	691	7	PP21-YEAST	SERINE/THREONINE PROT	3.51e+01
42	40	85.1	695	4	GFAL-SCHPO	POTATIVE GLUCOSAMINE-	3.51e+01
43	40	85.1	1020	2	CONT-MOUSE	CONTACTIN PRECURSOR	3.51e+01
44	40	85.1	1203	9	SDCI-CABEL	ZINC FINGER PROTEIN S	3.51e+01
45	40	85.1	1367	5	IG1R-HUMAN	INSULIN-LIKE GROWTH F	3.51e+01

#### ALIGNMENTS

RESULT 1  
 ID CALH\_MOUSE STANDARD; PRT; 1315 AA.  
 AC P39061;  
 DT 01-FEB-1995 (REL. 31, CREATED)  
 DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE COLLAGEN ALPHA 1(XVIII) CHAIN PRECURSOR.  
 GN COL18A1.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/C; TISSUE=LIVER;  
 RX MEDLINE; 94245707.  
 RA REHN M.V., HINTIKKA E., PIHLAJANIEMI T.;  
 RA J. BIOL. CHEM. 269:13929-13935(1994).  
 RN [2]  
 RP SEQUENCE OF 1-928 FROM N.A.  
 RX MEDLINE; 94240112.  
 RA REHN M.V., PIHLAJANIEMI T.;  
 RA PROC. NATL. ACAD. SCI. U.S.A. 91:4234-4238(1994).  
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- SIMILARITY: BELONGS TO THE MULTIPLEXIN FAMILY OF COLLAGENS.  
 DR ENBL; L16898; G553894; -  
 DR ENBL; U03714; G487734; -  
 DR ENBL; U03715; G1167905; -  
 DR ENBL; U34606; G1167905; JOINED.  
 DR ENBL; U34608; G1167905; JOINED.  
 DR ENBL; U34609; G1167905; JOINED.  
 DR ENBL; U34610; G1167905; JOINED.  
 DR ENBL; U34611; G1167905; JOINED.  
 DR ENBL; U34612; G1167905; JOINED.  
 DR ENBL; U34613; G1167905; JOINED.  
 DR ENBL; U03716; G1167905; JOINED.  
 DR ENBL; U03718; G1167905; JOINED.  
 KW EXTRACELLULAR MATRIX; CONNECTIVE TISSUE; REPEAT; HYDROXYLATION;  
 KW CELL ADHESION; COLLAGEN; GLYCOPROTEIN; SIGNAL.  
 FT SIGNAL 1 25  
 FT CHAIN 26 1315  
 FT COLLAGEN ALPHA 1(XVIII) CHAIN.  
 FT DOMAIN 26 326  
 FT NONHELICAL REGION 1 (NC1).  
 FT DOMAIN 327 353  
 FT TRIPLE-HELICAL REGION 1 (COL1).  
 FT DOMAIN 354 363  
 FT NONHELICAL REGION 2 (NC2).  
 FT DOMAIN 364 437  
 FT TRIPLE-HELICAL REGION 2 (COL2).  
 FT DOMAIN 438 461  
 FT NONHELICAL REGION 3 (NC3).

FT DOMAIN 462 583 TRIPLE-HELICAL REGION 3 (COL3).  
 FT DOMAIN 584 606 NONHELICAL REGION 4 (NC4).  
 FT DOMAIN 607 689 TRIPLE-HELICAL REGION 4 (COL4).  
 FT DOMAIN 690 703 NONHELICAL REGION 5 (NC5).  
 FT DOMAIN 704 745 TRIPLE-HELICAL REGION 5 (COL5).  
 FT DOMAIN 746 758 NONHELICAL REGION 6 (NC6).  
 FT DOMAIN 759 831 TRIPLE-HELICAL REGION 6 (COL6).  
 FT DOMAIN 832 841 NONHELICAL REGION 7 (NC7).  
 FT DOMAIN 842 874 TRIPLE-HELICAL REGION 7 (COL7).  
 FT DOMAIN 875 886 NONHELICAL REGION 8 (NC8).  
 FT DOMAIN 887 910 TRIPLE-HELICAL REGION 8 (COL8).  
 FT DOMAIN 911 917 NONHELICAL REGION 9 (NC9).  
 FT DOMAIN 918 969 TRIPLE-HELICAL REGION 9 (COL9).  
 FT DOMAIN 970 982 NONHELICAL REGION 10 (NC10).  
 FT DOMAIN 983 1000 TRIPLE-HELICAL REGION 10 (COL10).  
 FT DOMAIN 1001 1315 NONHELICAL REGION 11 (NC11).  
 FT CARBOHYD 126 126 POTENTIAL.  
 FT CARBOHYD 488 488 POTENTIAL.  
 SITE 892 894 CELL ATTACHMENT SITE (POTENTIAL).  
 SEQUENCE 1315 AA; 134263 MW; FA26C53 CRC32;

Query Match 100.0%; Score 47; DB 2; Length 1315;  
 Best Local Similarity 100.0%; Pred. No. 9.76e-01;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 74 yhpK 78  
 |||||  
 QY 2 YHFPK 6

RESULT 2  
 ID AACALSTAAU STANDARD; PRT; 479 AA.  
 AC P14507;  
 DT 01-JAN-1990 (REL. 13, CREATED)  
 DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)  
 DE 6'-AMINOGLYCOSIDE N-ACETYLTRANSFERASE (EC 2.3.1.1) (AAC(6')) / 2'-  
 DE 6'-AMINOGLYCOSIDE PHOSPHOTRANSFERASE (EC 2.7.1.1) (APH(2')).  
 GN AACAPHD.  
 OS STAPHYLOCOCCUS AUREUS, AND ENTEROCOCCUS FAECALIS (STREPTOCOCCUS  
 OS FAECALIS).  
 OC PROKARYOTA; FIRMICUTES; COCCI; MICROCOCCACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-S.AUREUS; TRANSPOSON-TN4001;  
 RA MEDLINE; 88187650.  
 ROUCH D.A., BYRNE M.E., KONG Y.C., SKURRAY R.A.;  
 J. GEN. MICROBIOL. 133:3039-3052(1987).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-E.FAECALIS;  
 RX MEDLINE; 86277923.  
 RA FERRETTI J.J., GILMORE K.S., COURVALIN P.;  
 J. BACTERIOL. 167:631-638(1986).  
 CC -!- FUNCTION: RESISTANCE TO GENTAMICIN, TOBRAMYCIN, AND KANAMYCIN.  
 CC TOBRAMYCIN AND KANAMYCIN RESISTANCE IS DUE TO THE ACC ACTIVITY,  
 CC SPECIFIED BY THE FIRST 170 AA, AND THE GENTAMICIN RESISTANCE IS  
 CC DUE TO THE APH ACTIVITY ENCODED BY THE C-TERMINAL REGION OF THE  
 CC PROTEIN.  
 CC -!- SIMILARITY: THE N-TERMINAL REGION IS HOMOLOGOUS TO STREPTOTHRICIN  
 CC ACETYLTRANSFERASE FROM S.LAVENDULAE. THE C-TERMINAL REGION IS  
 CC HOMOLOGOUS TO OTHERS APHS, AND TO A VIOMYCIN PHOSPHOTRANSFERASE  
 CC (VPH).  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 DR EMBL; M13771; G153586; -.  
 DR EMBL; M18086; G152948; -.  
 DR PIR; A26048; A26048.  
 DR PIR; S26353; S26353.  
 KW ANTIBIOTIC RESISTANCE; TRANSFERASE; KINASE; ATP-BINDING;  
 KW ACYLTRANSFERASE; TRANSPOSABLE ELEMENT.  
 FT DOMAIN 110 153 ACETYL-COA BINDING SITE (BY SIMILARITY).  
 FT SIMILAR 368 408 TO OTHERS APHS AND A VPH.  
 FT BINDING 393 393 AMINOGLYCOSIDE SUBSTRATES (BY

FT SQ SEQUENCE 479 AA; 56855 MW; A8193B91 CRC32; SIMILARITY).  
 Query Match 91.5%; Score 43; DB 1; Length 479;  
 Best Local Similarity 80.0%; Pred. No. 7.96e+00;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Db 86 yhpK 90  
 |||||  
 QY 2 YHFPK 6

RESULT 3  
 ID DPP6\_RAT STANDARD; PRT; 859 AA.  
 AC P46101;  
 DT 01-NOV-1995 (REL. 32, CREATED)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
 DE DIPEPTIDYL PEPTIDASE IV LIKE PROTEIN (DIPEPTIDYL AMINOPEPTIDASE-  
 DE RELATED PROTEIN) (DIPEPTIDYLPEPTIDASE VI) (DPPX-L/DPPX-S).  
 GN DPP6.  
 OS RATTUS NORVEGICUS (RAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RA MEDLINE; 92108018.  
 RA WADA K., YOKOTANI N., HUNTER C., DOI K., WENTHOLD R.J., SHIMASAKI S.;  
 PROC. NACL. ACAD. SCI. U.S.A. 89:197-201(1992).  
 CC -!- FUNCTION: MAY BE INVOLVED IN THE PHYSIOLOGICAL PROCESSES OF BRAIN  
 CC ACTIVITY MAY BE DUE TO THE SUBSTITUTION OF AN ASPARTATE RESIDUE  
 CC FOR THE SERINE RESIDUE IN THE PROPOSED CATALYTIC TRIAD.  
 CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).  
 CC -!- TISSUE SPECIFICITY: DPPX-S IS EXPRESSED IN BRAIN AND SOME  
 CC PERIPHERAL TISSUES INCLUDING KIDNEY, OVARY, AND TESTIS; IN  
 CC CONTRAST DPPX-L IS EXPRESSED ALMOST EXCLUSIVELY IN BRAIN.  
 CC -!- ALTERNATIVE PRODUCTS: TWO DISTINCT FORMS (DPPX-L AND -S) SHARE AN  
 CC IDENTICAL TRANSMEMBRANE DOMAIN AND A LONG C-TERMINAL EXTRACELLULAR  
 CC DOMAIN, BUT HAVE LONG AND SHORT N-TERMINAL CYTOPLASMIC DOMAINS.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B; ALSO KNOWN AS THE  
 CC PROLYL OLIGOPEPTIDASE FAMILY.  
 DR EMBL; M76426; G408714; -.  
 DR EMBL; M76427; G408716; -.  
 KW HYDROLASE; DIPEPTIDASE; SERINE PROTEASE; TRANSMEMBRANE; GLYCOPROTEIN;  
 KW SIGNAL-ANCHOR; ALTERNATIVE SPLICING.  
 FT DOMAIN 1 89 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 90 110 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT (POTENTIAL).  
 FT DOMAIN 111 859 EXTRACELLULAR (POTENTIAL).  
 FT CARBOHYD 167 167 POTENTIAL.  
 FT CARBOHYD 168 168 POTENTIAL.  
 FT CARBOHYD 313 313 POTENTIAL.  
 FT CARBOHYD 398 398 POTENTIAL.  
 FT CARBOHYD 465 465 POTENTIAL.  
 FT CARBOHYD 529 529 POTENTIAL.  
 FT CARBOHYD 560 560 POTENTIAL.  
 FT CARBOHYD 807 807 POTENTIAL.  
 FT VARSPLIC 1 75 MASLYORFTGKINTSRSPAPPASHLLGQGPEDAGSKP  
 FT LGPQAQAVPRGGAGGRFGYQARSDCDEED -> MT  
 FT AKPSAGSKVSQOQDQ (IN DPPX-S).  
 SQ SEQUENCE 859 AA; 97301 MW; B43COE82 CRC32;

Query Match 91.5%; Score 43; DB 3; Length 859;  
 Best Local Similarity 80.0%; Pred. No. 7.96e+00;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Db 336 yhpK 340  
 |||||  
 QY 2 YHFPK 6

RESULT 4  
 ID DPP6\_BOVIN STANDARD; PRT; 863 AA.  
 AC P42659;  
 DT 01-NOV-1995 (REL. 32, LAST CREATED)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
 DE DIPEPTIDYL PEPTIDASE IV LIKE PROTEIN (DIPEPTIDYL AMINOPEPTIDASE-RELATED PROTEIN) (DIPEPTIDYLPEPTIDASE VI) (DPPX-L/DPPX-S).  
 GN DPP6.  
 OS BOS TAURUS (BOVINE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; ARTIODACTYLA.  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE-BRAIN;  
 RX MEDLINE; 92108018  
 RA WADA K., YOKOTANI N., HUNTER C., DOI K., WENTHOLD R.J., SHIMASAKI S.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 89:197-201(1992).  
 CC -!- FUNCTION: MAY BE INVOLVED IN THE PHYSIOLOGICAL PROCESSES OF THE BRAIN  
 CC FUNCTION. HAS NO DIPEPTIDYL AMINOPEPTIDASE ACTIVITY. THE LACK OF  
 CC ACTIVITY MAY BE DUE TO THE SUBSTITUTION OF AN ASPARTATE RESIDUE  
 CC FOR THE SERINE RESIDUE IN THE PROPOSED CATALYTIC TRIAD.  
 CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).  
 CC -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN THE BRAIN. DPPX-L  
 CC IS EXPRESSED EXCLUSIVELY IN THE BRAIN WHEREAS DPPX-S IS FOUND IN  
 CC BRAIN, KIDNEY, OVARY AND TESTIS.  
 CC -!- ALTERNATIVE PRODUCTS: A SHORT FORM (DPPX-S) AND A LONG FORM  
 CC (DPPX-L) ARE PRODUCED BY ALTERNATIVE SPLICING. THE SEQUENCE  
 CC SHOWN HERE IS THAT OF THE LONG FORM (DPPX-L).  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B; ALSO KNOWN AS THE  
 CC PROLYL OLIGOPEPTIDASE FAMILY.  
 DR EMBL; M76428; G408718; -.  
 DR EMBL; M76429; G408720; -.  
 KW HYDROLASE; DIPEPTIDASE; SERINE PROTEASE; TRANSMEMBRANE; GLYCOPROTEIN;  
 KW SIGNAL-ANCHOR; ALTERNATIVE SPLICING.  
 FT DOMAIN 1 93 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 94 114 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT (POTENTIAL).  
 FT DOMAIN 115 863 EXTRACELLULAR (POTENTIAL).  
 FT CARBOHYD 171 171 POTENTIAL.  
 FT CARBOHYD 402 402 POTENTIAL.  
 FT CARBOHYD 469 469 POTENTIAL.  
 FT CARBOHYD 533 533 POTENTIAL.  
 FT CARBOHYD 564 564 POTENTIAL.  
 FT CARBOHYD 811 811 POTENTIAL.  
 FT VARSPLIC 1 79 MASLYORFTGKINTSRSPAPPEASRLGGOGPEEDGAGPK  
 FT PLGAQAPAAAPRRGGGGGAGGRPFQYQARSDDDED ->  
 FT MTAKEPNAGSKSVQOQEQ (IN DPPX-S).  
 SQ SEQUENCE 863 AA; 96556 MW; 8DA40472 CRC32;  
 Query Match 91.5%; Score 43; DB 3; Length 863;  
 Best Local Similarity 80.0%; Pred. No. 7.96e+00;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Db 340 yhyphk 344  
 11:11  
 QY 2 YHFPK 6

RESULT 5  
 ID DPP6\_HUMAN STANDARD; PRT; 865 AA.  
 AC P42658;  
 DT 01-NOV-1995 (REL. 32, CREATED)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
 DE DIPEPTIDYL PEPTIDASE IV LIKE PROTEIN (DIPEPTIDYL AMINOPEPTIDASE-RELATED PROTEIN) (DIPEPTIDYLPEPTIDASE VI) (DPPX-L/DPPX-S).  
 GN DPP6.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.

TISSUE-HIPPOCAMPUS;  
 RX MEDLINE; 93372805.  
 RA YOKOTANI N., DOI K., WENTHOLD R.J., WADA K.;  
 RL HUM. MOL. GENET. 2:1037-1039(1993).  
 CC -!- FUNCTION: MAY BE INVOLVED IN THE PHYSIOLOGICAL PROCESSES OF BRAIN  
 CC FUNCTION. HAS NO DIPEPTIDYL AMINOPEPTIDASE ACTIVITY. THE LACK OF  
 CC ACTIVITY MAY BE DUE TO THE SUBSTITUTION OF AN ASPARTATE RESIDUE  
 CC FOR THE SERINE RESIDUE IN THE PROPOSED CATALYTIC TRIAD.  
 CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).  
 CC -!- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN BRAIN.  
 CC -!- ALTERNATIVE PRODUCTS: A SHORT FORM (DPPX-S) AND A LONG FORM  
 CC (DPPX-L) ARE PRODUCED BY ALTERNATIVE SPLICING. THE SEQUENCE  
 CC SHOWN HERE IS THAT OF THE LONG FORM (DPPX-L).  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B; ALSO KNOWN AS THE  
 CC PROLYL OLIGOPEPTIDASE FAMILY.  
 DR EMBL; M96859; G306706; -.  
 DR EMBL; M96860; G306708; -.  
 DR MIM; 126141; -.  
 KW HYDROLASE; DIPEPTIDASE; SERINE PROTEASE; TRANSMEMBRANE; GLYCOPROTEIN;  
 KW SIGNAL-ANCHOR; ALTERNATIVE SPLICING.  
 FT DOMAIN 1 95 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 96 116 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT (POTENTIAL).  
 FT DOMAIN 117 865 EXTRACELLULAR (POTENTIAL).  
 FT CARBOHYD 173 173 POTENTIAL.  
 FT CARBOHYD 319 319 POTENTIAL.  
 FT CARBOHYD 404 404 POTENTIAL.  
 FT CARBOHYD 471 471 POTENTIAL.  
 FT CARBOHYD 535 535 POTENTIAL.  
 FT CARBOHYD 566 566 POTENTIAL.  
 FT CARBOHYD 813 813 POTENTIAL.  
 FT VARSPLIC 1 81 MASLYORFTGKINTSRSPAPPEASRLGGOGPEEDGGAGA  
 FT KPLGPRAQAAAPRRGGGGGAGGRPFQYQGRSDGDED  
 FT -> MTAKEPSAGSKSVQOQEQ (IN DPPX-S).  
 SQ SEQUENCE 865 AA; 97588 MW; 514C21E2 CRC32;  
 Query Match 91.5%; Score 43; DB 3; Length 865;  
 Best Local Similarity 80.0%; Pred. No. 7.96e+00;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Db 342 yhyphk 346  
 11:11  
 QY 2 YHFPK 6

RESULT 6  
 ID HEX2\_YEAST STANDARD; PRT; 1014 AA.  
 AC Q00816;  
 DT 01-APR-1993 (REL. 25, CREATED)  
 DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE HEX2 PROTEIN (SRN1 PROTEIN).  
 GN HEX2 OR SRN1 OR REG1 OR SPF43 OR YDR028C OR YD9813.06C.  
 OS SACCAROMYCES CEREVISIAE (BAKER'S YEAST).  
 OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 91364678.  
 RA NIEDERACHER D., ENTIAN K.-D.;  
 RL EUR. J. BIOCHEM. 200:311-319(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92269840.  
 RA TUNG K.-S., NORBECK L.L., NOLAN S.L., ATKINSON N.S., HOPPER A.K.;  
 RL MOL. CELL. BIOL. 12:2673-2680(1992).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C / AB972;  
 RA BOWMAN S., BARRELL B., RAJANDREAM M.A.;  
 RL SUBMITTED (JAN-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [4]  
 RP SEQUENCE OF 775-1014 FROM N.A.  
 RA EIDE L.G., SANDER C., PRYDZ H.;

RL YEAST 12:1085-1090(1996).  
 CC -!- FUNCTION: INVOLVED IN RNA PROCESSING AND NEGATIVE REGULATION OF  
 CC GLUCOSE REPRESSION. REGULATES THE LEVEL OF TWO ANTIGENS, P43 AND  
 CC P70.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL: M33703; G171666; -.  
 DR EMBL: M90540; G172699; -.  
 DR EMBL: 247814; G642300; -.  
 DR EMBL: X95966; E225541; -.  
 DR PIR: S17254; S17254.  
 DR PIR: S32613; S32613.  
 DR LISTA: SC00434; HEX.  
 DR SGD: L0001609; REG1.  
 KW NUCLEAR PROTEIN.  
 FT DOMAIN 277 283 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 543 552 SER-RICH.  
 FT DOMAIN 595 599 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 742 760 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 834 844 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 873 879 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT CONFLICT 376 377 DK -> EE (IN REF. 1).  
 FT CONFLICT 534 534 N -> K (IN REF. 1).  
 FT CONFLICT 657 657 D -> H (IN REF. 1).  
 FT CONFLICT 889 889 S -> T (IN REF. 1).  
 FT CONFLICT 988 1014 ARGMSKYLHSWKRSVDKPOENGNDSS -> QEVWQASTCT  
 FT (IN REF. 1).  
 FT LGKRVTSKKKMENTAVRRKNFVNMRK  
 FT (IN REF. 1).  
 SQ SEQUENCE 1014 AA; 112615 MW; 65BA11F7 CRC32;

Query Match 91.5%; Score 43; DB 4; Length 1014;  
 Best Local Similarity 80.0%; Pred. No. 7.96e+00;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 338 yhyfk 342  
 ||||  
 QY 2 YHFPK 6

RESULT 7  
 ID VNS2\_CVHOC STANDARD; PRT; 109 AA.  
 AC Q04853;  
 DT 01-FEB-1994 (REL. 28, CREATED)  
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)  
 DE NONSTRUCTURAL PROTEIN NS2 (NONSTRUCTURAL 12.9 KD PROTEIN).  
 OS HUMAN CORONAVIRUS (STRAIN OC43).  
 OS VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; CORONAVIRIDAE.  
 [1]  
 SEQUENCE FROM N.A.  
 RX MEDLINE: 93297129.  
 RA MOUNIR S., TALBOT P.J.;  
 RL VIROLOGY 192:355-360(1993).  
 DR EMBL: M99576; G329568; -.  
 DR PIR: A44275; A44275.  
 KW NONSTRUCTURAL PROTEIN.  
 SQ SEQUENCE 109 AA; 12935 MW; C47B6812 CRC32;

Query Match 89.4%; Score 42; DB 10; Length 109;  
 Best Local Similarity 80.0%; Pred. No. 1.32e+01;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 31 yqfpk 35  
 ||||  
 QY 2 YHFPK 6

RESULT 8  
 ID YJY9\_YEAST STANDARD; PRT; 349 AA.  
 AC P41903;  
 DT 01-NOV-1995 (REL. 32, CREATED)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE HYPOTHETICAL 40.3 KD PROTEIN IN ESS1-MER2 INTERGENIC REGION.

GN YJRO19C OR J1456.  
 OS SACHAROMYCES CREVISIAE (BAKER'S YEAST).  
 OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-DH484;  
 RA HANI J., STUMPF G., DOMDEY H.;  
 RL SUBMITTED (MAR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C / FY1679;  
 RA DE HAAN M., SMITS P.H.M., GRIVELL L.A.;  
 RL SUBMITTED (MAY-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: X85972; G758285; -.  
 DR EMBL: X87611; G854594; -.  
 DR EMBL: Z49519; G1015655; -.  
 KW HYPOTHETICAL PROTEIN.  
 SQ SEQUENCE 349 AA; 40259 MW; C38D06E2 CRC32;

Query Match 89.4%; Score 42; DB 11; Length 349;  
 Best Local Similarity 80.0%; Pred. No. 1.32e+01;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 191 yqfpk 195  
 ||||  
 QY 2 YHFPK 6

RESULT 9  
 ID YCR2\_BACTK STANDARD; PRT; 252 AA.  
 AC P21733;  
 DT 01-MAY-1991 (REL. 18, CREATED)  
 DT 01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)  
 DE HYPOTHETICAL 29.1 KD PROTEIN IN CRYB1 5'REGION (ORF2).  
 OS BACILLUS THURINGIENSIS (SUBSP. KURSTAKI).  
 OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-HD-1;  
 RX MEDLINE: 89123178.  
 RA WIDNER W.R., WHITELEY H.R.;  
 RL J. BACTERIOL. 171:965-974(1989).  
 DR EMBL: M23723; G1124821; -.  
 DR PIR: B32053; B32053.  
 KW HYPOTHETICAL PROTEIN; REPEAT.  
 FT DOMAIN 68 246 13 X 15 AA TANDEM REPEATS.  
 FT REPEAT 68 82 1.  
 FT REPEAT 83 97 2.  
 FT REPEAT 98 112 3.  
 FT REPEAT 113 127 4.  
 FT REPEAT 128 142 5.  
 FT REPEAT 143 157 6.  
 FT REPEAT 158 172 7.  
 FT REPEAT 173 187 8.  
 FT REPEAT 188 202 9.  
 FT REPEAT 203 217 10.  
 FT REPEAT 218 230 11 (INCOMPLETE).  
 FT REPEAT 231 239 12 (INCOMPLETE).  
 FT REPEAT 240 246 13 (INCOMPLETE).  
 SQ SEQUENCE 252 AA; 29110 MW; 2B6BECAB CRC32;

Query Match 87.2%; Score 41; DB 11; Length 252;  
 Best Local Similarity 80.0%; Pred. No. 2.16e+01;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 4 yhfpm 8  
 ||||  
 QY 2 YHFPK 6  
 RESULT 10  
 ID TRAL\_SALTY STANDARD; PRT; 101 AA.

```
AC P12058;
DT 01-OCT-1989 (REL. 12, LAST CREATED)
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE TRAL PROTEIN.
GN TRAL.
OS SALMONELLA TYPHIMURIUM.
OC PLASMID PED208.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC ENTEROBACTERIACEAE.
CC [1]
RN SEQUENCE FROM N.A.
RP MEDLINE; 87056998.
RX FINLAY B.B., FROST L.S., PARANCHYCH W.;
RA J. BACTERIOL. 168:990-998(1986).
CC -!- FUNCTION: MEMBRANE PROTEIN INVOLVED IN F PILIN FORMATION.
CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE (POTENTIAL).
CC -!- SIMILARITY: 44% SIMILARITY TO TRAL PROTEIN OF PLASMID F.
DR EMBL; M14733; G150665; -.
DR PIR; C25161; C25161.
OS PLASMID; CONJUGATION; OUTER MEMBRANE.
RN SEQUENCE 101 AA; 12037 MW; 941A2072 CRC32;
PRT; 255 AA.

Query Match 85.1%; Score 40; DB 9; Length 101;
Best Local Similarity 80.0%; Pred. No. 3.51e+01;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 9 yrfpk 13
QY 2 YHFPK 6

RESULT 11
ID YAL4_SCHPO STANDARD; PRT; 209 AA.
AC Q09927;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 23.8 KD PROTEIN C21E11.04 IN CHROMOSOME I.
GN SPAC21E11.04
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
CC [1]
RN SEQUENCE FROM N.A.
RP STRAIN-972;
RA MCLEAN J., HARRIS D., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- SIMILARITY: TO YEAST SPT10.
DR EMBL; Z67999; G1067220; -.
DE HYPOTHETICAL PROTEIN.
RN SEQUENCE 209 AA; 23809 MW; DA251012 CRC32;
PRT; 255 AA.

Query Match 85.1%; Score 40; DB 11; Length 209;
Best Local Similarity 100.0%; Pred. No. 3.51e+01;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 199 yhfpp 202
QY 2 YHFP 5

RESULT 12
ID VRPL_SALEN STANDARD; PRT; 255 AA.
AC P55219;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE 28.1 KD VIRULENCE PROTEIN.
GN SPVA.
OS SALMONELLA ENTERITIDIS.
OC PLASMID PNL2001
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC ENTEROBACTERIACEAE.
```

```
RN SEQUENCE FROM N.A.
RP STRAIN-AL1190;
RX MEDLINE; 94362897.
RA SUZUKI S., KOMASE K., MATSUI H., ABE A., KAWAHARA K., TAMURA Y.,
RA KIJIMA M., DANBARA H., NAKAMURA M., SATO S.;
RL MICROBIOLOGY 140:1307-1318(1994).
CC -!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
CC OF SALMONELLAS.
CC -!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
CC PLASMIDS.
DR EMBL; D14490; G517163; -.
KW PLASMID; VIRULENCE.
SQ SEQUENCE 255 AA; 28200 MW; EAE5ED6D CRC32;
PRT; 255 AA.

Query Match 85.1%; Score 40; DB 10; Length 255;
Best Local Similarity 100.0%; Pred. No. 3.51e+01;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 135 yhfpp 138
QY 2 YHFP 5

RESULT 13
ID VRPL_SALDU STANDARD; PRT; 255 AA.
AC P24418;
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT 01-MAR-1992 (REL. 21, LAST ANNOTATION UPDATE)
DE 28.1 KD VIRULENCE PROTEIN.
GN VSDB.
OS SALMONELLA DUBLIN.
OC PLASMID PSDL2.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC ENTEROBACTERIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91251759.
RA KRAUSE M., ROUDIER C., FIERER J., HARWOOD J., GUINEY D.;
RL MOL. MICROBIOL. 5:307-316(1991).
CC -!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
CC OF SALMONELLAS.
CC -!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
CC PLASMIDS.
DR EMBL; X56727; G47838; -.
DR PIR; S15214; S15214.
KW PLASMID; VIRULENCE.
SQ SEQUENCE 255 AA; 28156 MW; 2455A4CB CRC32;
PRT; 255 AA.

Query Match 85.1%; Score 40; DB 10; Length 255;
Best Local Similarity 100.0%; Pred. No. 3.51e+01;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 135 yhfpp 138
QY 2 YHFP 5

RESULT 14
ID VRPL_SALCH STANDARD; PRT; 255 AA.
AC P17449;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
DE 28.1 KD VIRULENCE PROTEIN (PROTEIN M3, IN MBA REGION).
OS SALMONELLA CHOLERAE-SUIS (SALMONELLA ENTERICA).
OC PLASMID PKDC50.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC ENTEROBACTERIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RF-1;
```

RX MEDLINE; 90245675.  
RA MATSUI H.;  
RL NUCLEIC ACIDS RES. 18:2181-2181(1990).  
CC -!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE  
CC OF SALMONELLAS.  
CC -!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE  
CC PLASMIDS.  
DR EMBL; X52035; G46897; -.  
DR PIR; S09497; S09497.  
KW PLASMID: VIRULENCE.  
SQ SEQUENCE 255 AA; 28184 MW; ECBB81FC CRC32;  
  
Query Match 85.1%; Score 40; DB 10; Length 255;  
Best Local Similarity 100.0%; Pred. No. 3.51e+01;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Db 135 yhf 138  
|||  
2 YHFP 5  
  
RESULT 15  
ID COX3 ANOGA STANDARD; PRT; 262 AA.  
AC P34842;  
DT 01-FEB-1994 (REL. 28, CREATED)  
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)  
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
DE CYTOCHROME C OXIDASE POLYPEPTIDE III (EC 1.9.3.1).  
GN COIII.  
OS ANOPHELES GAMBIAE (AFRICAN MALARIA MOSQUITO).  
OG MITOCHONDRION.  
OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-G3;  
RA BEARD C.B.; HAMM D.M.; COLLINS F.H.;  
RL INSECT MOL. BIOL. 2:103-104(1993).  
CC -!- FUNCTION: SUBUNIT I, II, AND III FORM THE FUNCTIONAL CORE OF  
CC THE ENZYME COMPLEX.  
CC -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +  
CC 4 FERRICYTOCHROME C.  
DR EMBL; L20934; G309062; -.  
KW OXIDOREDUCTASE; MITOCHONDRION; TRANSMEMBRANE.  
SQ SEQUENCE 262 AA; 30132 MW; 73CE1E48 CRC32;  
  
Query Match 85.1%; Score 40; DB 2; Length 262;  
Best Local Similarity 80.0%; Pred. No. 3.51e+01;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
Db 226 yhf 230  
|||  
2 YHFP 6

Search completed: Tue Dec 2 14:42:53 1997  
Job time : 8 secs.

\*\*\*\*\*

WQSRFL (TM)

\*\*\*\*\*

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by Intelligent, Inc.

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Dec 3 10:06:49 1997; Master time 5.21 Seconds  
33.068 Million cell updates/sec

ular output not generated.

Title: >US-08-915-004-2  
Description: (1-14) from US08915004.pap  
Perfect Score: 78  
Sequence: 1 QHSHXQFOTFLXK 14

Scoring table: PAM 150  
Gap 15

Searched: 101610 seqs, 12294212 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq28  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21

Statistics: Mean 16.442; Variance 48.941; scale 0.336

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Full No.	Score	Query Match	Length	ID	Description	Pred. No.
1	78	100.0	14	20	R99922 Osteoclastogenesis in	2.43e-02
2	78	100.0	272	20	R99944 Mutated OCIF, OCIF-CD	2.43e-02
3	78	100.0	321	20	R99949 Mutated OCIF, OCIF-CS	2.43e-02
4	78	100.0	327	20	R99941 Mutated OCIF, OCIF-DD	2.43e-02
5	78	100.0	351	20	R99943 Mutated OCIF, OCIF-CC	2.43e-02
6	78	100.0	359	20	R99939 Mutated OCIF, OCIF-DC	2.43e-02
7	78	100.0	359	20	R99937 Mutated OCIF, OCIF-DC	2.43e-02
8	78	100.0	360	20	R99938 Mutated OCIF, OCIF-DC	2.43e-02
9	78	100.0	360	20	R99936 Mutated OCIF, OCIF-DC	2.43e-02
10	78	100.0	380	20	R99932 Mature osteoclastogen	2.43e-02
11	78	100.0	390	20	R99357 Human tumour necrosis	2.43e-02
12	78	100.0	393	20	R99948 Mutated OCIF, OCIF-CB	2.43e-02
13	78	100.0	399	20	R99942 Mutated OCIF, OCIF-CL	2.43e-02
14	78	100.0	401	20	R99934 Mutated OCIF, OCIF-C2	2.43e-02
15	78	100.0	401	20	R99932 Mutated OCIF, OCIF-C2	2.43e-02
16	78	100.0	401	20	R99935 Mutated OCIF, OCIF-C2	2.43e-02
17	78	100.0	401	20	R99933 Mutated OCIF, OCIF-C2	2.43e-02
18	78	100.0	401	20	R99931 Mutated OCIF, OCIF-C1	2.43e-02
19	78	100.0	401	20	R99925 Full length osteoclas	2.43e-02
20	46	59.0	22	18	W02241 Prostate-specific mem	1.19e+02

21	46	59.0	22	10	R55102	Prostate-specific mem	1.19e+02
22	46	59.0	494	13	R72367	Human auxillary cytoc	1.19e+02
23	46	59.0	494	13	R72368	Human auxillary cytoc	1.19e+02
24	46	59.0	494	18	R93174	Human cytochrome P450	1.19e+02
25	46	59.0	494	17	R81466	Human derived cytochr	1.19e+02
26	46	59.0	494	18	R93175	Human cytochrome P450	1.19e+02
27	46	59.0	506	9	R50010	Secretory alkaline ph	1.19e+02
28	46	59.0	531	1	P91776	Germ cell alkaline ph	1.19e+02
29	46	59.0	532	4	R20527	Human germ cell alka	1.19e+02
30	46	59.0	730	18	W02234	Prostate-specific mem	1.19e+02
31	46	59.0	730	10	R55097	Prostate-specific mem	1.19e+02
32	45	57.7	227	19	W05530	Mouse TRADD intracell	1.52e+02
33	45	57.7	312	19	W05528	Human TRADD intracell	1.52e+02
34	45	57.7	425	3	R13792	E75B exon B1 polypept	1.52e+02
35	45	57.7	432	14	R78274	Chicken TEF-1A.	1.52e+02
36	45	57.7	445	14	R78275	Chicken TEF-1B.	1.52e+02
37	45	57.7	559	3	R13263	JMI-229 cell line t-p	1.52e+02
38	45	57.7	2308	10	R52580	RPTP-beta amino acid	1.52e+02
39	45	57.7	2308	11	R57902	Human RPTP-beta.	1.52e+02
40	44	56.4	189	3	P50328	Bovine interferon alp	1.94e+02
41	44	56.4	189	3	P30074	Sequence of bovine le	1.94e+02
42	44	56.4	189	3	P50325	Bovine interferon alp	1.94e+02
43	44	56.4	358	2	R05123	Bat-PA(L).	1.94e+02
44	44	56.4	431	2	R06459	v-PA.Beta.	1.94e+02
45	44	56.4	477	2	R05122	Bat-PA(H).	1.94e+02

#### ALIGNMENTS

RESULT 1  
ID R99922 standard; peptide; 14 AA.  
AC R99922;  
DT 22-APR-1997 (first entry)  
DE Osteoclastogenesis inhibitory factor peptide fragment #2.  
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
KW osteoporosis.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT Misc\_difference 1  
FT /note= "Any amino acid"  
FT /note= "Any amino acid"  
FT /note= "Any amino acid"  
FT /note= "Any amino acid"  
FT /note= "Any amino acid"  
FT /note= "Any amino acid"  
PD WO9626217-A1.  
PN 29-AUG-1996.  
PF 20-FEB-1996; J00374.  
PR 20-FEB-1995; JP-054977.  
PR 21-JUL-1995; JP-207508.  
PA (SNOW) SNOW BRAND MILK PROD CO LTD.  
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
DR WPI; 96-402320/40.  
PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
PT for bone resorption control, esp. treatment of osteoporosis  
PS Claim 1; Page 61; 183pp; Japanese.  
CC The sequences given in R99921-23 and R99926 represent fragments of  
CC the osteoclastogenesis inhibitory factor (OCIF) of the invention.  
CC The OCIF has a molecular weight by SDS-PAGE of 60 kD under reducing  
CC conditions and 120 kD under non-reducing conditions. The protein is  
CC adsorbed onto cation-exchangers or heparin and its activity is  
CC lowered after 10 mins at 70 deg.C or 30 mins at 56 deg.C, and is  
CC lost after 10 mins at 90 deg.C. OCIF is useful in the control of  
CC bone resorption and therefore in the treatment and prevention of  
CC disorders of bone resorption, e.g. osteoporosis.  
SQ Sequence 14 AA;

Query Match 100.0%; Score 78; DB 20; Length 14;  
Best Local Similarity 100.0%; Pred. No. 2.43e-02;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2 qhsxqetfqlxk 14  
|||||

QY 2 QHSXQEQTFQLXK 14

RESULT 2

ID R99944 standard; Protein; 272 AA.

AC R99944;

DT 23-APR-1997 (first entry)

DE Mutated OCIF, OCIF-CDD2.

KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;

OS Synthetic.

EH Key Location/Qualifiers

FT Peptide 1..21

FT /note= "Signal peptide"

FT Protein 22..272

FT /note= "Mature OCIF-CDD2"

PN WO9626217-A1.

PD 29-AUG-1996.

PF 20-FEB-1996; J00374.

PR 20-FEB-1995; JP-054977.

PR 21-JUL-1995; JP-207508.

PA (SNOW ) SNOW BRAND MILK PROD CO LTD.

PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;

PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;

DR WPI: 96-402320/40.

DR N-PSDB; T33174.

PT DNA encoding osteoclastogenesis inhibitory factor protein - useful

PT for bone resorption control, esp. treatment of osteoporosis

PS Claim 68; Page 121-122; 183pp; Japanese.

CC This sequence represents a mutated version of the full length

CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This

CC sequence represents OCIF-CDD2 in which amino acids 252-380 of the

CC mature OCIF protein are deleted. The OCIF of the invention

CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions

CC and 120 kD under non-reducing conditions. The protein is adsorbed onto

CC cation-exchangers or heparin and its activity is lowered after 10 mins

CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90

CC deg.C. OCIF is useful in the control of bone resorption and therefore

CC in the treatment and prevention of disorders of bone resorption, e.g.

CC osteoporosis.

SQ Sequence 272 AA;

Query Match 100.0%; Score 78; DB 20; Length 272;

Best Local Similarity 84.6%; Pred. No. 2.43e-02;

Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

243 qhssqeqtfqlk 255

|||||

2 QHSXQEQTFQLXK 14

RESULT 3

ID R99949 standard; Protein; 321 AA.

AC R99949;

DT 23-APR-1997 (first entry)

DE Mutated OCIF, OCIF-Csph.

KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;

OS Synthetic.

EH Key Location/Qualifiers

FT Peptide 1..21

FT /note= "Signal peptide"

FT Protein 22..321

FT /note= "Mature OCIF-Csph"

PN WO9626217-A1.

PD 29-AUG-1996.

PF 20-FEB-1996; JP-054977.

PR 20-FEB-1995; JP-207508.

PR 21-JUL-1995; JP-207508.

PA (SNOW ) SNOW BRAND MILK PROD CO LTD.

PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;

PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;

DR WPI: 96-402320/40.

DR N-PSDB; T33179.

PT DNA encoding osteoclastogenesis inhibitory factor protein - useful

PT for bone resorption control, esp. treatment of osteoporosis

PS Claim 83; Page 128-129; 183pp; Japanese.

CC This sequence represents a mutated version of the full length

CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This

CC sequence represents OCIF-Csph in which amino acids 298-380 of the mature

CC OCIF protein are replaced by Ser-Leu-Asp. These changes are caused by

CC the introduction of a restriction site in the DNA encoding this protein.

CC The OCIF of the invention has a molecular weight by SDS-PAGE of 60 kD

CC under reducing conditions and 120 kD under non-reducing conditions. The

CC protein is adsorbed onto cation-exchangers or heparin and its activity is

CC lowered after 10 mins at 70 deg.C or 30 mins at 56 deg.C, and is lost

CC after 10 mins at 90 deg.C. OCIF is useful in the control of bone

CC resorption and therefore in the treatment and prevention of disorders

CC of bone resorption, e.g. osteoporosis.

SQ Sequence 321 AA;

Query Match 100.0%; Score 78; DB 20; Length 321;

Best Local Similarity 84.6%; Pred. No. 2.43e-02;

Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 243 qhssqeqtfqlk 255

|||||

2 QHSXQEQTFQLXK 14

QY 2 QHSXQEQTFQLXK 14

RESULT 4

ID R99941 standard; Protein; 327 AA.

AC R99941;

DT 23-APR-1997 (first entry)

DE Mutated OCIF, OCIF-DDD2.

KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;

OS Synthetic.

EH Key Location/Qualifiers

FT Peptide 1..21

FT /note= "Signal peptide"

FT Protein 22..327

FT /note= "Mature OCIF-DDD2"

FT Misc difference 273..274

FT /note= "Position of deletion, delta 253-326"

PN WO9626217-A1.

PD 29-AUG-1996.

PF 20-FEB-1996; J00374.

PR 20-FEB-1995; JP-054977.

PR 21-JUL-1995; JP-207508.

PA (SNOW ) SNOW BRAND MILK PROD CO LTD.

PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;

PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;

DR WPI: 96-402320/40.

DR N-PSDB; T33171.

PT DNA encoding osteoclastogenesis inhibitory factor protein - useful

PT for bone resorption control, esp. treatment of osteoporosis

PS Claim 59; Page 115-116; 183pp; Japanese.

CC This sequence represents a mutated version of the full length

CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This

CC sequence represents OCIF-DDD2 in which amino acids 253-326 of the

CC mature OCIF protein are deleted. The OCIF of the invention

CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions

CC and 120 kD under non-reducing conditions. The protein is adsorbed onto

CC cation-exchangers or heparin and its activity is lowered after 10 mins

CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90

CC deg.C. OCIF is useful in the control of bone resorption and therefore

CC in the treatment and prevention of disorders of bone resorption, e.g.

CC osteoporosis.

SQ Sequence 327 AA;

Query Match 100.0%; Score 78; DB 20; Length 327;

Best Local Similarity 84.6%; Pred. No. 2.43e-02;

Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 243 qhssqeqtfqlk 255



```

QY      2 QHSXQEQTFOLXK 14
      III IIIIIII I
RESULT  5
ID R99943 standard; Protein: 351 AA.
AC R99943;
DT 23-APR-1997 (first entry)
DE Mutated OCIF, OCIF-CC.
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;
OS Osteoporosis.
FH Key Location/Qualifiers
FT Peptide 1..21
FT /note= "Signal peptide"
FT Protein 22..351
FT /note= "Mature OCIF-CC"
PN W09626217-A1.
PD 29-AUG-1996.
PF 20-FEB-1996; J00374.
PR 21-JUL-1995; JP-207508.
PA (SNOW ) SNOW BRAND MILK PROD CO LTD.
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;
DR WPI: 96-402320/40.
DR N-PSDB; T33169.
DR DNA encoding osteoclastogenesis inhibitory factor protein - useful
PT for bone resorption control, esp. treatment of osteoporosis
PS Claim 65; Page 119-121; 183pp; Japanese.
CC This sequence represents a mutated version of the full length
CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This
CC sequence represents OCIF-CC in which amino acids 331-380 of the
CC mature OCIF protein are deleted. The OCIF of the invention
CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions
CC and 120 kD under non-reducing conditions. The protein is adsorbed onto
CC cation-exchangers or heparin and its activity is lowered after 10 mins
CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90
CC deg.C. OCIF is useful in the control of bone resorption and therefore
CC in the treatment and prevention of disorders of bone resorption, e.g.
CC osteoporosis.
SQ Sequence 351 AA;
      Query Match 100.0%; Score 78; DB 20; Length 351;
      Best Local Similarity 84.6%; Pred. No. 2.43e-02;
      Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 243 qhssqeqtfqlk 255
      III IIIIIII I
      2 QHSXQEQTFOLXK 14
RESULT  6
ID R99939 standard; Protein: 359 AA.
AC R99939;
DT 23-APR-1997 (first entry)
DE Mutated OCIF, OCIF-DCR4
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;
OS Osteoporosis.
FH Key Location/Qualifiers
FT Peptide 1..21
FT /note= "Signal peptide"
FT Protein 22..359
FT /note= "Mature OCIF-DCR4"
FT Misc.difference 143..144
FT /note= "Position of deletion, delta 123-164"
PN W09626217-A1.
PD 29-AUG-1996.
PF 20-FEB-1996; J00374.
PR 20-FEB-1995; JP-054977.
PR 21-JUL-1995; JP-207508.
PA (SNOW ) SNOW BRAND MILK PROD CO LTD.
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;
DR WPI: 96-402320/40.
DR N-PSDB; T33167.
DR DNA encoding osteoclastogenesis inhibitory factor protein - useful
PT for bone resorption control, esp. treatment of osteoporosis
PS Claim 47; Page 107-109; 183pp; Japanese.
CC This sequence represents a mutated version of the full length
CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This
CC sequence represents OCIF-DCR2 in which amino acids 43-84 of the
CC mature OCIF protein are deleted. The OCIF of the invention
CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions
CC and 120 kD under non-reducing conditions. The protein is adsorbed onto
CC cation-exchangers or heparin and its activity is lowered after 10 mins
CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90
CC deg.C. OCIF is useful in the control of bone resorption and therefore
CC in the treatment and prevention of disorders of bone resorption, e.g.
CC osteoporosis.
SQ Sequence 359 AA;
      Query Match 100.0%; Score 78; DB 20; Length 351;
      Best Local Similarity 84.6%; Pred. No. 2.43e-02;
      Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 201 qhssqeqtfqlk 213
      III IIIIIII I
      2 QHSXQEQTFOLXK 14
RESULT  7
ID R99937 standard; Protein: 359 AA.
AC R99937;
DT 23-APR-1997 (first entry)
DE Mutated OCIF, OCIF-DCR2.
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;
OS Osteoporosis.
FH Key Location/Qualifiers
FT Peptide 1..21
FT /note= "Signal peptide"
FT Protein 22..359
FT /note= "Mature OCIF-DCR2"
FT Misc.difference 63..64
FT /note= "Position of deletion, delta 43-84"
PN W09626217-A1.
PD 29-AUG-1996.
PF 20-FEB-1996; J00374.
PR 20-FEB-1995; JP-054977.
PR 21-JUL-1995; JP-207508.
PA (SNOW ) SNOW BRAND MILK PROD CO LTD.
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;
DR WPI: 96-402320/40.
DR N-PSDB; T33167.
DR DNA encoding osteoclastogenesis inhibitory factor protein - useful
PT for bone resorption control, esp. treatment of osteoporosis
PS Claim 47; Page 107-109; 183pp; Japanese.
CC This sequence represents a mutated version of the full length
CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This
CC sequence represents OCIF-DCR2 in which amino acids 43-84 of the
CC mature OCIF protein are deleted. The OCIF of the invention
CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions
CC and 120 kD under non-reducing conditions. The protein is adsorbed onto
CC cation-exchangers or heparin and its activity is lowered after 10 mins
CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90
CC deg.C. OCIF is useful in the control of bone resorption and therefore
CC in the treatment and prevention of disorders of bone resorption, e.g.
CC osteoporosis.
SQ Sequence 359 AA;
      Query Match 100.0%; Score 78; DB 20; Length 359;
      Best Local Similarity 84.6%; Pred. No. 2.43e-02;
      Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 QHSXQEQTFOLXK 14
      III IIIIIII I
      2 QHSXQEQTFOLXK 14

```

```

Db 201 qhsseqtfqlk 213
QY 2 QHSXQEQTFQLXK 14

RESULT 8
ID R9938 standard; Protein: 360 AA.
AC R9938;
DT 23-APR-1997 (first entry)
DE Mutated OCIF, OCIF-PCR3
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;
OS Synthetic.
FH Key Location/Qualifiers
FT Peptide 1..21
FT /note= "Signal peptide"
FT Protein 22..360
FT /note= "Mature OCIF-PCR3"
FT Misc_difference 105..106
FT /note= "Position of deletion, delta 85-122"
PN WO9626217-A1.
PD 29-AUG-1996.
PF 20-FEB-1996; J00374.
PR 20-FEB-1995; JP-054977.
PR 21-JUL-1995; JP-207508.
PA (SNOW ) SNOW BRAND MILK PROD CO LTD.
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;
DR WPI: 96-402320/40.
DR N-PSDB; T33168.
PT DNA encoding osteoclastogenesis inhibitory factor protein - useful
PT for bone resorption control, esp. treatment of osteoporosis
PS Claim 4; Page 105-107; 183pp; Japanese.
CC This sequence represents a mutated version of the full length
CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This
CC sequence represents OCIF-PCR1 in which amino acids 2-42 of the
CC mature OCIF protein are deleted. The OCIF of the invention
CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions
CC and 120 kD under non-reducing conditions. The protein is adsorbed onto
CC cation-exchangers or heparin and its activity is lowered after 10 mins
CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90
CC deg.C. OCIF is useful in the control of bone resorption and therefore
CC in the treatment and prevention of disorders of bone resorption, e.g.
CC osteoporosis.
SQ Sequence 360 AA;

Query Match 100.0%; Score 78; DB 20; Length 360;
Best Local Similarity 84.6%; Pred. No. 2.43e-02;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 202 qhsseqtfqlk 214
QY 2 QHSXQEQTFQLXK 14

RESULT 10
ID R9924 standard; Protein: 380 AA.
AC R9924;
DT 22-APR-1997 (first entry)
DE Mature osteoclastogenesis inhibitory factor.
DE Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;
KW Osteoporosis.
OS Homo sapiens.
PN WO9626217-A1.
PD 29-AUG-1996.
PF 20-FEB-1996; J00374.
PR 20-FEB-1995; JP-054977.
PR 21-JUL-1995; JP-207508.
PA (SNOW ) SNOW BRAND MILK PROD CO LTD.
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;
DR WPI: 96-402320/40.
DR N-PSDB; T36685.
PT DNA encoding osteoclastogenesis inhibitory factor protein - useful
PT for bone resorption control, esp. treatment of osteoporosis
PS Claim 6; Page 62-64; 183pp; Japanese.
CC This sequence represents the mature osteoclastogenesis inhibitory
CC factor (OCIF) of the invention. The OCIF has a molecular weight by
CC SDS-PAGE of 60 kD under reducing conditions and 120 kD under non-
CC reducing conditions. The protein is adsorbed onto cation-exchangers
CC or heparin and its activity is lowered after 10 mins at 70 deg.C or
CC 30 mins at 56 deg.C, and is lost after 10 mins at 90 deg.C. OCIF is
CC useful in the control of bone resorption and therefore in the
CC treatment and prevention of disorders of bone resorption, e.g.
CC osteoporosis.
SQ Sequence 380 AA;

Query Match 100.0%; Score 78; DB 20; Length 380;
Best Local Similarity 84.6%; Pred. No. 2.43e-02;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 222 qhsseqtfqlk 234
QY 2 QHSXQEQTFQLXK 14

RESULT 10
ID R9924 standard; Protein: 380 AA.
AC R9924;
DT 22-APR-1997 (first entry)
DE Mature osteoclastogenesis inhibitory factor.
DE Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;
KW Osteoporosis.
OS Homo sapiens.
PN WO9626217-A1.
PD 29-AUG-1996.
PF 20-FEB-1996; J00374.
PR 20-FEB-1995; JP-054977.
PR 21-JUL-1995; JP-207508.
PA (SNOW ) SNOW BRAND MILK PROD CO LTD.
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;
DR WPI: 96-402320/40.
DR N-PSDB; T36685.
PT DNA encoding osteoclastogenesis inhibitory factor protein - useful
PT for bone resorption control, esp. treatment of osteoporosis
PS Claim 6; Page 62-64; 183pp; Japanese.
CC This sequence represents the mature osteoclastogenesis inhibitory
CC factor (OCIF) of the invention. The OCIF has a molecular weight by
CC SDS-PAGE of 60 kD under reducing conditions and 120 kD under non-
CC reducing conditions. The protein is adsorbed onto cation-exchangers
CC or heparin and its activity is lowered after 10 mins at 70 deg.C or
CC 30 mins at 56 deg.C, and is lost after 10 mins at 90 deg.C. OCIF is
CC useful in the control of bone resorption and therefore in the
CC treatment and prevention of disorders of bone resorption, e.g.
CC osteoporosis.
SQ Sequence 380 AA;

Query Match 100.0%; Score 78; DB 20; Length 380;
Best Local Similarity 84.6%; Pred. No. 2.43e-02;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 222 qhsseqtfqlk 234
QY 2 QHSXQEQTFQLXK 14

RESULT 10
ID R9924 standard; Protein: 380 AA.
AC R9924;
DT 22-APR-1997 (first entry)
DE Mature osteoclastogenesis inhibitory factor.
DE Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;
KW Osteoporosis.
OS Homo sapiens.
PN WO9626217-A1.
PD 29-AUG-1996.
PF 20-FEB-1996; J00374.
PR 20-FEB-1995; JP-054977.
PR 21-JUL-1995; JP-207508.
PA (SNOW ) SNOW BRAND MILK PROD CO LTD.
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;
DR WPI: 96-402320/40.
DR N-PSDB; T36685.
PT DNA encoding osteoclastogenesis inhibitory factor protein - useful
PT for bone resorption control, esp. treatment of osteoporosis
PS Claim 6; Page 62-64; 183pp; Japanese.
CC This sequence represents the mature osteoclastogenesis inhibitory
CC factor (OCIF) of the invention. The OCIF has a molecular weight by
CC SDS-PAGE of 60 kD under reducing conditions and 120 kD under non-
CC reducing conditions. The protein is adsorbed onto cation-exchangers
CC or heparin and its activity is lowered after 10 mins at 70 deg.C or
CC 30 mins at 56 deg.C, and is lost after 10 mins at 90 deg.C. OCIF is
CC useful in the control of bone resorption and therefore in the
CC treatment and prevention of disorders of bone resorption, e.g.
CC osteoporosis.
SQ Sequence 380 AA;

Query Match 100.0%; Score 78; DB 20; Length 380;
Best Local Similarity 84.6%; Pred. No. 2.43e-02;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 205 qhsseqtfqlk 217
QY 2 QHSXQEQTFQLXK 14

RESULT 9
ID R9936 standard; Protein: 360 AA.
AC R9936;
DT 23-APR-1997 (first entry)
DE Mutated OCIF, OCIF-PCR1.
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;
OS Synthetic.
FH Key Location/Qualifiers
FT Peptide 1..21
FT /note= "Signal peptide"
FT Protein 22..360
FT /note= "Mature OCIF-PCR1"
FT Misc_difference 22..23
FT /note= "Position of deletion, delta 2-42"
PN WO9626217-A1.
PD 29-AUG-1996.

```

```

PS Claim 80; Page 126-128; 183pp; Japanese.
CC This sequence represents a mutated version of the full length
CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This
CC sequence represents OCIF-CBst in which Gln 371 is substituted by
CC Leu and amino acids 373-380 of the mature OCIF protein are deleted.
CC These changes are caused by the introduction of a restriction site in
CC the DNA encoding this protein. The OCIF of the invention has a
CC molecular weight by SDS-PAGE of 60 kD under reducing conditions
CC and 120 kD under non-reducing conditions. The protein is adsorbed onto
CC cation-exchangers or heparin and its activity is lowered after 10 mins
CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90
CC deg.C. OCIF is useful in the control of bone resorption and therefore
CC in the treatment and prevention of disorders of bone resorption, e.g.
CC osteoporosis.
SQ Sequence 393 AA;

Query Match 100.0%; Score 78; DB 20; Length 393;
Best Local Similarity 84.6%; Pred. No. 2.43e-02;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 243 qhssgeqtfqlk 255
QY 2 QHSXQEQTFQLXK 14

RESULT 13
ID R99942 standard; Protein; 399 AA.
AC R99942;
DT 23-APR-1997 (first entry)
DE Mutated OCIF, OCIF-CL.
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;
KW osteoporosis.
OS Synthetic.
FH Key Location/Qualifiers
FT Peptide 1..21
FT /note= "Signal peptide"
FT Protein 22..399
FT /note= "Mature OCIF-CL"
PN WO9626217-A1.
PD 29-AUG-1996; J00374.
PF 20-FEB-1995; JP-054977.
PR 20-FEB-1995; JP-054977.
PR 21-JUL-1995; JP-207508.
PA (SNOW) SNOW BRAND MILK PROD CO LTD.
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;
DR WPI: 96-402320/40.
DR N-PSDB: T33172.
PT DNA encoding osteoclastogenesis inhibitory factor protein - useful
PT for bone resorption control, esp. treatment of osteoporosis
PS Claim 62; Page 117-119; 183pp; Japanese.
CC This sequence represents a mutated version of the full length
CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This
CC sequence represents OCIF-CL in which amino acids 379-380 of the
CC mature OCIF protein are deleted. The OCIF of the invention
CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions
CC and 120 kD under non-reducing conditions. The protein is adsorbed onto
CC cation-exchangers or heparin and its activity is lowered after 10 mins
CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90
CC deg.C. OCIF is useful in the control of bone resorption and therefore
CC in the treatment and prevention of disorders of bone resorption, e.g.
CC osteoporosis.
SQ Sequence 399 AA;

Query Match 100.0%; Score 78; DB 20; Length 399;
Best Local Similarity 84.6%; Pred. No. 2.43e-02;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 243 qhssgeqtfqlk 255
QY 2 QHSXQEQTFQLXK 14

PS Claim 80; Page 126-128; 183pp; Japanese.
CC This sequence represents a mutated version of the full length
CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This
CC sequence represents OCIF-CBst in which Gln 371 is substituted by
CC Leu and amino acids 373-380 of the mature OCIF protein are deleted.
CC These changes are caused by the introduction of a restriction site in
CC the DNA encoding this protein. The OCIF of the invention has a
CC molecular weight by SDS-PAGE of 60 kD under reducing conditions
CC and 120 kD under non-reducing conditions. The protein is adsorbed onto
CC cation-exchangers or heparin and its activity is lowered after 10 mins
CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90
CC deg.C. OCIF is useful in the control of bone resorption and therefore
CC in the treatment and prevention of disorders of bone resorption, e.g.
CC osteoporosis.
SQ Sequence 393 AA;

Query Match 100.0%; Score 78; DB 20; Length 393;
Best Local Similarity 84.6%; Pred. No. 2.43e-02;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 243 qhssgeqtfqlk 255
QY 2 QHSXQEQTFQLXK 14

RESULT 12
ID R99948 standard; Protein; 393 AA.
AC R99948;
DT 23-APR-1997 (first entry)
DE Mutated OCIF, OCIF-CBst.
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;
KW osteoporosis.
OS Synthetic.
FH Key Location/Qualifiers
FT Peptide 1..21
FT /note= "Signal peptide"
FT Protein 22..393
FT /note= "Mature OCIF-CBst"
FT Misc_difference 392
FT /label= Gln371Leu
PN WO9626217-A1.
PD 29-AUG-1996.
PF 20-FEB-1996; J00374.
PR 20-FEB-1995; JP-054977.
PR 21-JUL-1995; JP-207508.
PA (SNOW) SNOW BRAND MILK PROD CO LTD.
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;
DR WPI: 96-402320/40.
DR N-PSDB: T33178.
PT DNA encoding osteoclastogenesis inhibitory factor protein - useful
PT for bone resorption control, esp. treatment of osteoporosis

```

```

RESULT 14
ID R99934 standard; Protein; 401 AA.
AC R99934;
DT 22-APR-1997 (first entry)
DE Mutated OCIF, OCIF-C22S.
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;
OS Osteoporosis.
FH Key Location/Qualifiers
FT Peptide 1..21
FT /note= "Signal peptide"
FT Protein 22..401
FT /note= "Mature OCIF-C22S"
FT Misc_difference 277
FT /label= C22S
PN WO9626217-A1.
PD 29-AUG-1996.
PR 20-FEB-1996; J00374.
PR 21-JUL-1995; JP-207508.
PA (SNOW ) SNOW BRAND MILK PROD CO LTD.
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;
DR WPI; 96-402320/40.
DR N-PSDB; T33164.
PT DNA encoding osteoclastogenesis inhibitory factor protein - useful
PT for bone resorption control, esp. treatment of osteoporosis
PS Claim 38; Page 100-102; 183pp; Japanese.
CC This sequence represents a mutated version of the full length
CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This
CC sequence represents OCIF-C22S in which the 22nd Cys residue in the
CC mature OCIF protein is substituted by Ser. The OCIF of the invention
CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions
CC and 120 kD under non-reducing conditions. The protein is adsorbed onto
CC cation-exchangers or heparin and its activity is lowered after 10 mins
CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90
CC deg.C. OCIF is useful in the control of bone resorption and therefore
CC in the treatment and prevention of disorders of bone resorption, e.g.
CC osteoporosis.
SQ Sequence 401 AA;

Query Match 100.0%; Score 78; DB 20; Length 401;
Best Local Similarity 84.6%; Pred. No. 2.43e-02;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 243 qhsseqtfqlk 255
||| |||||
2 QHSXQEQTFQLXK 14

RESULT 15
ID R99932 standard; Protein; 401 AA.
AC R99932;
DT 22-APR-1997 (first entry)
DE Mutated OCIF, OCIF-C20S.
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;
OS Osteoporosis.
FH Key Location/Qualifiers
FT Peptide 1..21
FT /note= "Signal peptide"
FT Protein 22..401
FT /note= "Mature OCIF-C20S"
FT Misc_difference 202
FT /label= C20S
PN WO9626217-A1.
PD 29-AUG-1996.
PR 20-FEB-1996; J00374.
PR 21-JUL-1995; JP-207508.
PA (SNOW ) SNOW BRAND MILK PROD CO LTD.
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;
DR WPI; 96-402320/40.
DR N-PSDB; T33165.
PT DNA encoding osteoclastogenesis inhibitory factor protein - useful
PT for bone resorption control, esp. treatment of osteoporosis
PS Claim 41; Page 103-105; 183pp; Japanese.
CC This sequence represents a mutated version of the full length
CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This
CC sequence represents OCIF-C23S in which the 23rd Cys residue in the
CC mature OCIF protein is substituted by Ser. The OCIF of the invention
CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions
CC and 120 kD under non-reducing conditions. The protein is adsorbed onto
CC cation-exchangers or heparin and its activity is lowered after 10 mins
CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90
CC deg.C. OCIF is useful in the control of bone resorption and therefore
CC in the treatment and prevention of disorders of bone resorption, e.g.
CC osteoporosis.
SQ Sequence 401 AA;

Query Match 100.0%; Score 78; DB 20; Length 401;
Best Local Similarity 84.6%; Pred. No. 2.43e-02;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 243 qhsseqtfqlk 255
||| |||||
2 QHSXQEQTFQLXK 14

RESULT 15
ID R99932 standard; Protein; 401 AA.
AC R99932;
DT 22-APR-1997 (first entry)
DE Mutated OCIF, OCIF-C20S.
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;
OS Osteoporosis.
FH Key Location/Qualifiers
FT Peptide 1..21
FT /note= "Signal peptide"
FT Protein 22..401
FT /note= "Mature OCIF-C20S"
FT Misc_difference 202
FT /label= C20S
PN WO9626217-A1.
PD 29-AUG-1996.
PR 20-FEB-1996; J00374.
PR 21-JUL-1995; JP-207508.
PA (SNOW ) SNOW BRAND MILK PROD CO LTD.
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;
DR WPI; 96-402320/40.
DR N-PSDB; T33165.
PT DNA encoding osteoclastogenesis inhibitory factor protein - useful
PT for bone resorption control, esp. treatment of osteoporosis
PS Claim 41; Page 103-105; 183pp; Japanese.
CC This sequence represents a mutated version of the full length
CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This
CC sequence represents OCIF-C23S in which the 23rd Cys residue in the
CC mature OCIF protein is substituted by Ser. The OCIF of the invention
CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions
CC and 120 kD under non-reducing conditions. The protein is adsorbed onto
CC cation-exchangers or heparin and its activity is lowered after 10 mins
CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90
CC deg.C. OCIF is useful in the control of bone resorption and therefore
CC in the treatment and prevention of disorders of bone resorption, e.g.
CC osteoporosis.
SQ Sequence 401 AA;

Query Match 100.0%; Score 78; DB 20; Length 401;
Best Local Similarity 84.6%; Pred. No. 2.43e-02;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 243 qhsseqtfqlk 255
||| |||||
2 QHSXQEQTFQLXK 14

```

```

DR WPI; 96-402320/40.
DR N-PSDB; T33162.
PT DNA encoding osteoclastogenesis inhibitory factor protein - useful
PT for bone resorption control, esp. treatment of osteoporosis
PS Claim 32; Page 96-98; 183pp; Japanese.
CC This sequence represents a mutated version of the full length
CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This
CC sequence represents OCIF-C20S in which the 20th Cys residue in the
CC mature OCIF protein is substituted by Ser. The OCIF of the invention
CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions
CC and 120 kD under non-reducing conditions. The protein is adsorbed onto
CC cation-exchangers or heparin and its activity is lowered after 10 mins
CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90
CC deg.C. OCIF is useful in the control of bone resorption and therefore
CC in the treatment and prevention of disorders of bone resorption, e.g.
CC osteoporosis.
SQ Sequence 401 AA;

Query Match 100.0%; Score 78; DB 20; Length 401;
Best Local Similarity 84.6%; Pred. No. 2.43e-02;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 243 qhsseqtfqlk 255
||| |||||
2 QHSXQEQTFQLXK 14

RESULT 16
ID R99935 standard; Protein; 401 AA.
AC R99935;
DT 22-APR-1997 (first entry)
DE Mutated OCIF, OCIF-C23S.
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;
OS Osteoporosis.
FH Key Location/Qualifiers
FT Peptide 1..21
FT /note= "Signal peptide"
FT Protein 22..401
FT /note= "Mature OCIF-C23S"
FT Misc_difference 400
FT /label= C23S
PN WO9626217-A1.
PD 29-AUG-1996.
PR 20-FEB-1996; JP-054977.
PR 21-JUL-1995; JP-207508.
PA (SNOW ) SNOW BRAND MILK PROD CO LTD.
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;
DR WPI; 96-402320/40.
DR N-PSDB; T33165.
PT DNA encoding osteoclastogenesis inhibitory factor protein - useful
PT for bone resorption control, esp. treatment of osteoporosis
PS Claim 41; Page 103-105; 183pp; Japanese.
CC This sequence represents a mutated version of the full length
CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This
CC sequence represents OCIF-C23S in which the 23rd Cys residue in the
CC mature OCIF protein is substituted by Ser. The OCIF of the invention
CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions
CC and 120 kD under non-reducing conditions. The protein is adsorbed onto
CC cation-exchangers or heparin and its activity is lowered after 10 mins
CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90
CC deg.C. OCIF is useful in the control of bone resorption and therefore
CC in the treatment and prevention of disorders of bone resorption, e.g.
CC osteoporosis.
SQ Sequence 401 AA;

Query Match 100.0%; Score 78; DB 20; Length 401;
Best Local Similarity 84.6%; Pred. No. 2.43e-02;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 243 qhsseqtfqlk 255
||| |||||
2 QHSXQEQTFQLXK 14

```

```

QY      III IIIIIII I
        2 QHSXQEQTFOLXK 14

RESULT 17
ID R99933 standard; Protein: 401 AA.
AC R99933; 1997 (first entry)
DT 22-APR-1997 (first entry)
DE Mutated OCIF, OCIF-C21S.
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;
OS osteoporosis.
FH Synthetic.
FH Key Location/Qualifiers
FT Peptide 1..21
FT /note= "Signal peptide"
FT Protein 22..401
FT /note= "Mature OCIF-C21S"
FT Misc-difference 277
FT /label= C21S
PN WO9626217-A1.
PD 29-AUG-1996.
PD 20-FEB-1996; J00374.
PD 20-FEB-1995; JP-054977.
PD 21-JUL-1995; JP-207508.
PA (SNOW ) SNOW BRAND MILK PROD CO LTD.
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;
DR WPI: 96-402320/40.
DR N-PSDB: T33163.
PT DNA encoding osteoclastogenesis inhibitory factor protein - useful
PT for bone resorption control, esp. treatment of osteoporosis
PS Claim 35; Page 98-100; 183pp; Japanese.
CC This sequence represents a mutated version of the full length
CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This
CC sequence represents OCIF-C21S in which the 21st Cys residue in the
CC mature OCIF protein is substituted by Ser. The OCIF of the invention
CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions
CC and 120 kD under non-reducing conditions. The protein is adsorbed onto
CC cation-exchangers or heparin and its activity is lowered after 10 mins
CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90
CC deg.C. OCIF is useful in the control of bone resorption and therefore
CC in the treatment and prevention of disorders of bone resorption, e.g.
CC osteoporosis.
SQ Sequence 401 AA;

Query Match 100.0%; Score 78; DB 20; Length 401;
Best Local Similarity 84.6%; Pred. No. 2.43e-02;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 243 qhssqeqtfqlk 255
QY      III IIIIIII I
        2 QHSXQEQTFOLXK 14

RESULT 19
ID R99925 standard; Protein: 401 AA.
AC R99925; 1997 (first entry)
DT 22-APR-1997 (first entry)
DE Full length osteoclastogenesis inhibitory factor.
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;
OS Osteoporosis.
FH Key Location/Qualifiers
FT Peptide 1..21
FT /note= "Signal peptide"
FT Protein 22..401
FT /note= "Mature OCIF, claim 6"
PN WO9626217-A1.
PD 29-AUG-1996.
PD 20-FEB-1996; J00374.
PD 20-FEB-1995; JP-054977.
PD 21-JUL-1995; JP-207508.
PA (SNOW ) SNOW BRAND MILK PROD CO LTD.
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;
DR WPI: 96-402320/40.
DR N-PSDB: T36685.
PT DNA encoding osteoclastogenesis inhibitory factor protein - useful
PT for bone resorption control, esp. treatment of osteoporosis
PS Disclosure; Page 64-66; 183pp; Japanese.
CC This sequence represents the full length osteoclastogenesis inhibitory
CC factor (OCIF) of the invention. The OCIF has a molecular weight by
CC SDS-PAGE of 60 kD under reducing conditions and 120 kD under non-
CC reducing conditions. The protein is adsorbed onto cation-exchangers
CC or heparin and its activity is lowered after 10 mins at 70 deg.C or
CC 30 mins at 56 deg.C, and is lost after 10 mins at 90 deg.C. OCIF is
CC useful in the control of bone resorption and therefore in the
CC treatment and prevention of disorders of bone resorption, e.g.
CC osteoporosis.
SQ Sequence 401 AA;

Query Match 100.0%; Score 78; DB 20; Length 401;
Best Local Similarity 84.6%; Pred. No. 2.43e-02;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 243 qhssqeqtfqlk 255

```

```

PR 21-JUL-1995; JP-207508.
PA (SNOW ) SNOW BRAND MILK PROD CO LTD.
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;
DR WPI: 96-402320/40.
DR N-PSDB: T33161.
PT DNA encoding osteoclastogenesis inhibitory factor protein - useful
PT for bone resorption control, esp. treatment of osteoporosis
PS Claim 29; Page 94-96; 183pp; Japanese.
CC This sequence represents a mutated version of the full length
CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This
CC sequence represents OCIF-C19S in which the 19th Cys residue in the
CC mature OCIF protein is substituted by Ser. The OCIF of the invention
CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions
CC and 120 kD under non-reducing conditions. The protein is adsorbed onto
CC cation-exchangers or heparin and its activity is lowered after 10 mins
CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90
CC deg.C. OCIF is useful in the control of bone resorption and therefore
CC in the treatment and prevention of disorders of bone resorption, e.g.
CC osteoporosis.
SQ Sequence 401 AA;

Query Match 100.0%; Score 78; DB 20; Length 401;
Best Local Similarity 84.6%; Pred. No. 2.43e-02;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 243 qhssqeqtfqlk 255
QY      III IIIIIII I
        2 QHSXQEQTFOLXK 14

RESULT 19
ID R99925 standard; Protein: 401 AA.
AC R99925; 1997 (first entry)
DT 22-APR-1997 (first entry)
DE Full length osteoclastogenesis inhibitory factor.
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;
OS Osteoporosis.
FH Key Location/Qualifiers
FT Peptide 1..21
FT /note= "Signal peptide"
FT Protein 22..401
FT /note= "Mature OCIF, claim 6"
PN WO9626217-A1.
PD 29-AUG-1996.
PD 20-FEB-1996; J00374.
PD 20-FEB-1995; JP-054977.
PD 21-JUL-1995; JP-207508.
PA (SNOW ) SNOW BRAND MILK PROD CO LTD.
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;
DR WPI: 96-402320/40.
DR N-PSDB: T36685.
PT DNA encoding osteoclastogenesis inhibitory factor protein - useful
PT for bone resorption control, esp. treatment of osteoporosis
PS Disclosure; Page 64-66; 183pp; Japanese.
CC This sequence represents the full length osteoclastogenesis inhibitory
CC factor (OCIF) of the invention. The OCIF has a molecular weight by
CC SDS-PAGE of 60 kD under reducing conditions and 120 kD under non-
CC reducing conditions. The protein is adsorbed onto cation-exchangers
CC or heparin and its activity is lowered after 10 mins at 70 deg.C or
CC 30 mins at 56 deg.C, and is lost after 10 mins at 90 deg.C. OCIF is
CC useful in the control of bone resorption and therefore in the
CC treatment and prevention of disorders of bone resorption, e.g.
CC osteoporosis.
SQ Sequence 401 AA;

Query Match 100.0%; Score 78; DB 20; Length 401;
Best Local Similarity 84.6%; Pred. No. 2.43e-02;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 243 qhssqeqtfqlk 255

```

```

CC The inventors attempted to sequence the PSM Ag by modified Edman
CC degradation. Peptides (R55098-107) that gave clear dominant peaks
CC on HPLC from the digested PSM Ag sample were used. Some of the
CC residues were unidentified, and others were present at very low
CC levels and identified with lower confidence. Some of these peptides
CC were used to design primers to carry out PCR to identify cDNA clones
CC encoding the PSM Ag. R55102 was used to design primers Q65527-30.
CC A clone, IN-20 was identified as a partial PSM sequence.
CC (The full PSM Ag sequence is shown in R55097 and is encoded
CC by Q65520). The PSM coding sequence is useful for suppressing or
CC modulating the metastatic ability of prostate tumour cells to grow,
CC or for eliminating them. The protein is useful to identify or purify
CC ligands of the Ag. It is also an attractive target for Ab-directed
CC imaging and targeting of prostatic tumour deposits.
CC Sequence 22 AA;
SQ

```

Query Match 59.0%; Score 46; DB 10; Length 22;  
 Best Local Similarity 75.0%; Pred. No. 1.19e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

Db 15 eqnfglak 22
QY 7 EQTFOLXK 14

```

RESULT 22

```

ID R72367 standard; Protein; 494 AA.
AC R72367;
DT 14-NOV-1995 (first entry)
DE Human auxillary cytochrome P450 species 2A6 protein.
KW Human cytochrome P450; amplification; PCR; primer; expression vector;
KW yeast NADPH-P450 reductase; safety; fusion protein; metabolite;
OS Homo sapiens.
PN EP-644267-A.
PD 22-MAR-1995.
PF 20-JUL-1994; 111298.
PR 20-JUL-1993; JP-201120.
PR 21-JUL-1993; JP-180246.
PR 30-JUL-1993; JP-208279.
PA (HAYA/) HAYASHI K.
PA (SUMO) SUMITOMO CHEM CO LTD.
PI Hayashi K, Kaneko H, Komai K, Nakatsuka I, Sakaki T;
PI Yabusaki Y;
DR WPI; 95-116991/16.
DR N-PSDB; Q87721.
PT Evaluation of safety of a chemical cpd. - using recombinant yeast
PT expressing human cytochrome p450 and a yeast NADPH-P450 reductase
PS Examples; Page 49-51; 124pp; English.
CC The amino acid sequence of the human auxillary cytochrome P450 species
CC 2A6. The cDNA was amplified by PCR using the primers Q87751-4. The
CC product was cloned into the yeast expression vectors pAAH5N or pAHRH to
CC produce the vectors p2A6 for the expression of the cytochrome P450 alone
CC or p2A6R for co-expression with the yeast NADPH-P450 reductase.
CC The vectors are used in a method for evaluating the safety of a chemical
CC compound by reacting the chemical compound with recombinantly produced
CC human cytochrome P450 molecular species 1A2 (Q87714), 2C9 (Q87715), 2E1
CC (Q87716), or 3A4 (Q87718), or their auxillary species and variants
CC (Q87718-32), and yeast NADPH-P450 reductase, either as a fused protein or
CC in cell extracts, and analysing the resulting metabolite to assess the
CC safety of the chemical compound. The method is useful for determining
CC whether the chemical compound, or its metabolite, will be converted into
CC a carcinogenic or mutagenic form through metabolism in the liver.
CC Sequence 494 AA;
SQ

```

Query Match 59.0%; Score 46; DB 13; Length 494;  
 Best Local Similarity 55.6%; Pred. No. 1.19e+02;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

```

Db 232 gpqgqafql 240
QY 4 SXEQTFQL 12

```

RESULT 23  
ID R72368 standard; Protein: 494 AA.  
AC R72368;  
DT 14-NOV-1995 (first entry)  
DE Human auxillary cytochrome P450 species 2A6 variant 1 protein.  
KW Human cytochrome P450; amplification; PCR; primer; expression vector;  
KW yeast NADPH-P450 reductase; safety; fusion protein; metabolite;  
KW carcinogen; mutagen; liver metabolism.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT Misc.difference 476  
FT /note="Arg to Lys variation"  
FN EP-644267-A.  
PD 22-MAR-1995.  
PF 20-JUL-1994; 111298.  
PR 20-JUL-1993; JP-201120.  
PR 21-JUL-1993; JP-180246.  
PR 30-JUL-1993; JP-208279.  
PA (HAYA/) HAYASHI K.  
PA (SUMO ) SUMITOMO CHEM CO LTD.  
PA Hayashi K, Kaneko H, Komai K, Nakatsuka I, Sakaki T;  
Yabusaki Y;  
WPI: 95-116991/16.  
N-PSDB; Q87722.  
PT Evaluation of safety of a chemical cpd. - using recombinant yeast  
PT expressing human cytochrome p450 and a yeast NADPH-P450 reductase  
PS Examples; Page 53-55; 124pp; English.  
CC The amino acid sequence of the human auxillary cytochrome P450 species  
CC 2A6 variant 1. This variant contains a variation at residue 476: Arg to  
CC Lys, caused by a variation at base 1427: G to A in the DNA sequence. The  
CC cDNA was amplified by PCR using the primers Q87751-4. The product was  
CC cloned into the yeast expression vectors pAAH5N or pAHRH to produce the  
CC vectors p2A6 variant 1 for the expression of the cytochrome P450 alone or  
CC p2A6R variant 1 for co-expression with the yeast NADPH-P450 reductase.  
CC The vectors are used in a method for evaluating the safety of a chemical  
CC compound by reacting the chemical compound with recombinantly produced  
CC human cytochrome P450 molecular species 1A2 (Q87714), 2C9 (Q87715), 2E1  
CC (Q87716), or 3A4 (Q87718), or their auxillary species and variants  
CC (Q87718-32), and yeast NADPH-P450 reductase, either as a fused protein or  
CC in cell extracts, and analysing the resulting metabolite to assess the  
CC safety of the chemical compound. The method is useful for determining  
CC whether the chemical compound, or its metabolite, will be converted into  
CC a carcinogenic or mutagenic form through metabolism in the liver.  
CC Sequence 494 AA;  
Query Match 59.0%; Score 46; DB 13; Length 494;  
Best Local Similarity 55.6%; Pred. No. 1.19e+02;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
232 gpqgqafql 240  
:|:|:|  
QY 4 SXEQTFQL 12

RESULT 24  
ID R93174 standard; Protein: 494 AA.  
AC R93174;  
DT 11-OCT-1996 (first entry)  
DE Human cytochrome P450 molecular species 2A6 variant #1 protein.  
KW Human cytochrome P450; amplified; PCR; polymerase chain reaction; primer;  
KW liver; yeast; expression vector; NADPH-P450 reductase; ADH gene promoter;  
KW evaluation; safety; fusion protein; metabolite; detoxification;  
KW carcinogenic.  
OS Homo sapiens.  
PN J08056695-A.  
PD 05-MAR-1996.  
PF 15-JUL-1994; 164184.  
PR 20-JUL-1993; JP-201120.  
PR 30-JUL-1993; JP-208279.  
PR 17-JUN-1994; JP-136053.  
PA (SUMO ) SUMITOMO CHEM CO LTD.  
PA WPI: 96-182311/19.

DR N-PSDB; T28387.  
PT Novel method for the evaluation of the safety of a cpd. - using a  
PT human cytochrome P450 and yeast NADPH reductase to determine whether  
PT the analyte cpd. is detoxified or metabolised to a carcinogen  
PS Example 1; Page 33-35; 74pp; Japanese.  
CC This is the amino acid sequence of the human cytochrome P450 molecular  
CC species 2A6 variant #1 protein. The corresp. gene was amplified from a  
CC human liver derived cDNA library as 2 fragments of 0.6 and 0.9 kb using  
CC primers T26941-44. The prod. was cloned into the yeast expression vector  
CC pAAH5N to generate plasmid p2A6 for prodn. of the cytochrome only or into  
CC the vector pAHRH to generate the plasmid p2A6R for co-prodn. with the  
CC yeast NADPH-P450 reductase. The sequence is placed under control of the  
CC yeast ADH gene promoter and terminator.  
CC The vectors are used in a method for evaluating the safety of a cpd. by  
CC reacting the test cpd. with recombinantly produced human cytochrome P450  
CC mol. species 1A2 (T28380), 2C9 (T28381), 2E1 (T28382), 3A4 (T28383) or  
CC their variants (T28384-98) together with yeast NADPH-P450 reductase  
CC (either as a fused protein or as a cell extract) and analysing the  
CC resultant metabolite. The cpd. is considered "safe" if it is detoxified  
CC or not rendered carcinogenic or "unsafe" if it is not detoxified or is  
CC metabolised to a carcinogenic cpd.  
CC Sequence 494 AA;  
Query Match 59.0%; Score 46; DB 18; Length 494;  
Best Local Similarity 55.6%; Pred. No. 1.19e+02;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
232 gpqgqafql 240  
:|:|:|  
QY 4 SXEQTFQL 12

RESULT 25  
ID R81466 standard; Protein: 494 AA.  
AC R81466;  
DT 01-AUG-1996 (first entry)  
DE Human derived cytochrome P4502A6.  
KW Human derived cytochrome; P4502A6; commercial cDNA library; yeast;  
KW transfection; recombinant production; expression vector; mammal;  
KW immunisation; sensitisation; antibody; determination; detection;  
KW non-cross reactive.  
OS Homo sapiens.  
PN J08027195-A.  
PD 30-JAN-1996.  
PF 13-JUL-1994; JP-161550.  
PR 13-JUL-1994; JP-161550.  
PA (SUMO ) SUMITOMO CHEM CO LTD.  
DR WPI: 96-136336/14.  
DR N-PSDB; T17409.  
PT Antibody recognising human derived cytochrome P4502A6 - allows  
PT specific detection of cytochrome P450 species in humans  
PS Example 1; Pages 11-13; 13pp; Japanese.  
CC The present sequence is the human derived cytochrome (HDC)  
CC P4502A6, which was obtd. from a commercial cDNA library. Yeast  
CC were transfected with an expression vector contg. the HDC cDNA, the  
CC cultured and then disrupted to give a microsomal fraction. The  
CC HDC was purified from the fraction, and used to immunise and  
CC sensitise a mammal. Blood was drawn from the mammal, and an  
CC anti-HDC antibody isolated. The antibody obtd. recognises HDC  
CC P4502A6 partic. at a serum dilution rate of 1:10000, and is  
CC substantially without cross reaction to other HDC P450 spp..  
CC Sequence 494 AA;  
Query Match 59.0%; Score 46; DB 17; Length 494;  
Best Local Similarity 55.6%; Pred. No. 1.19e+02;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
232 gpqgqafql 240  
:|:|:|  
QY 4 SXEQTFQL 12

RESULT 26

R93175 standard; Protein; 494 AA.  
AC R93175;  
DI 11-Oct-1996 (first entry)  
DE Human cytochrome P450 molecular species 2A6 variant #2 protein.  
KW Human cytochrome P450; amplified; PCR; polymerase chain reaction; primer;  
KW liver; yeast; expression vector; NADPH-P450 reductase; ADH gene promoter;  
KW evaluation; safety; fusion protein; metabolite; detoxification;  
KW carcinogenic.  
OS Homo sapiens.  
PN J08056695-A.  
PD 05-MAR-1996.  
PF 15-JUL-1994; 164184.  
PR 20-JUL-1993; JP-201120.  
PR 30-JUL-1993; JP-208279.  
PR 17-JUN-1994; JP-136053.  
PA (SUMO) SUMITOMO CHEM CO LTD.  
DR WPI; 96-182311/19.  
N-PSDB; T28388.  
Novel method for the evaluation of the safety of a cpd. - using a  
human cytochrome P450 and yeast NADPH reductase to determine whether  
the analyte cpd. is detoxified or metabolised to a carcinogen  
Example 1; Page 35-37; 74pp; Japanese.  
CC This is the amino acid sequence of the human cytochrome P450 molecular  
species 2A6 variant #2 protein. The corresp. gene was amplified from a  
human liver derived cDNA library as 2 fragments of 0.6 and 0.9 kb using  
primers T26941-44. The prod. was cloned into the yeast expression vector  
pAAH5N to generate plasmid p2A6 for prodn. of the cytochrome only or into  
the vector pAHR to generate the plasmid p2A6R for co-prodn. with the  
yeast NADPH-P450 reductase. The sequence is placed under control of the  
yeast ADH gene promoter and terminator.  
CC The vectors are used in a method for evaluating the safety of a cpd. by  
reacting the test cpd. with recombinantly produced human cytochrome P450  
mol. species 1A2 (T28380), 2C9 (T28381), 2E1 (T28382), 3A4 (T28383) or  
their variants (T28384-98) together with yeast NADPH-P450 reductase  
CC (either as a fused protein or as a cell extract) and analysing the  
CC resultant metabolite. The cpd. is considered "safe" if it is detoxified  
CC or not rendered carcinogenic or "unsafe" if it is not detoxified or is  
CC metabolised to a carcinogenic cpd.  
SQ Sequence 494 AA;

Query Match 59.0%; Score 46; DB 18; Length 494;

Best Local Similarity 55.6%; Pred. NO. 1.19e+02;

Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 232 gpqgqafql 240

:|||

4 SXQEQFQL 12

Search completed: Wed Dec 3 10:06:57 1997

Job time : 8 secs.



\*\*\*\*\*

WQSERLH (TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (C) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Tue Dec 2 14:44:39 1997; Maspar time 2.99 Seconds  
135.144 Million cell updates/sec  
Circular output not generated.

Title: >US-08-915-004-2  
Description: (1-14) from US08915004.pep  
Perfect Score: 78  
Sequence: 1 QHXSXQEQTFQLXK 14

Scoring table: PAM 150  
Gap 15

Searched: 91006 seqs, 2888923 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: p1r51  
1:ann1 2:ann2 3:ann3 4:ann4 5:annn1 6:annn2 7:unann3  
8:unann4 9:unann5 10:unann6 11:unann7 12:unann8  
13:unann9 14:unann10 15:unenc 16:unrev

Statistics: Mean 22.756; Variance 33.196; scale 0.686

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Query No.	Score	Match	Length	ID	Description	Pred. No.
1	55	70.5	585	13	gene H4(D10S170) pro	1.44e+00
2	53	67.9	634	4	W1WL51	3.38e+00
3	53	67.9	638	8	S36546	3.38e+00
4	53	67.9	1893	12	A56158	3.38e+00
5	52	66.7	113	12	S43435	5.14e+00
6	52	66.7	113	12	S32723	5.14e+00
7	52	66.7	864	12	A49070	5.14e+00
8	51	65.4	450	3	A43733	7.78e+00
9	51	65.4	631	8	S36505	7.78e+00
10	50	64.1	1458	7	A49707	1.17e+01
11	49	62.8	245	6	B33956	1.75e+01
12	49	62.8	506	9	H64096	1.75e+01
13	48	61.5	139	14	S36325	2.61e+01
14	48	61.5	387	3	A43704	2.61e+01
15	48	61.5	467	3	A49377	2.61e+01
16	48	61.5	629	4	W1WL31	2.61e+01
17	48	61.5	755	9	B41836	2.61e+01
18	48	61.5	757	9	I40923	2.61e+01
19	48	61.5	786	8	S37031	2.61e+01
20	48	61.5	786	8	B49349	2.61e+01
21	47	60.3	241	11	S58635	3.87e+01

22 47 60.3 335 8 A49897 anthranilate phospho 3.87e+01  
23 47 60.3 337 16 S66515 protein kinase (EC 2 3.87e+01  
24 47 60.3 426 14 A54882 transcriptional enha 3.87e+01  
25 47 60.3 426 14 S40779 transcription factor 3.87e+01  
26 47 60.3 426 13 A40032 transcription enhanc 3.87e+01  
27 47 60.3 430 14 S41767 transcription enhanc 3.87e+01  
28 47 60.3 566 11 S53813 probable RNA helicas 3.87e+01  
29 47 60.3 566 11 S49262 RNA helicase - slime 3.87e+01  
30 47 60.3 570 12 S24459 hypothetical protein 3.87e+01  
31 46 59.0 134 16 S41538 aroQ protein - Actin 5.69e+01  
32 46 59.0 189 6 S23709 interferon alpha-1 p 5.69e+01  
33 46 59.0 275 2 TVRTFR transforming protein 5.69e+01  
34 46 59.0 331 1 O4HUPB cytochrome P450 2A3, 5.69e+01  
35 46 59.0 494 1 O4HUA6 coumarin 7-hydroxyla 5.69e+01  
36 46 59.0 532 5 S12076 alkaline phosphatase 5.69e+01  
37 46 59.0 532 13 I37991 alkaline phosphatase 5.69e+01  
38 46 59.0 535 1 PAHUA alkaline phosphatase 5.69e+01  
39 46 59.0 644 4 W1WL58 E1 protein - human p 5.69e+01  
40 46 59.0 713 11 S56833 nuclear pore protein 5.69e+01  
41 46 59.0 750 13 A56881 prostate-specific me 5.69e+01  
42 46 59.0 1248 10 A47445 reverse gyrase - Sul 5.69e+01  
43 46 59.0 1511 11 A53151 pleiotropic drug res 5.69e+01  
44 46 59.0 1529 11 S69688 hypothetical protein 5.69e+01  
45 45 57.7 559 5 A29941 t-plasminogen activa 8.33e+01

ALIGNMENTS

RESULT 1  
ENTRY #type complete  
TITLE gene H4(D10S170) protein - human  
ORGANISM #formal\_name Homo sapiens #common\_name man  
DATE 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 02-Jul-1996  
ACCESSIONS I58403  
REFERENCE I58403  
#authors Grieco, M.; Cerrato, A.; Santoro, M.; Fusco, A.; Melillo, R.M.; Vecchio, G.  
#journal Oncogene (1994) 9:2531-2535  
#title Cloning and characterization of H4 (D10S170), a gene involved in RET rearrangements in vivo.  
#accession I58403  
#status preliminary; translated from GB/EMBL/DBDJ  
#molecule\_type mRNA  
#residues 1-585 #label RES  
#cross-references GB:S72869; NID:g633869; CDS\_PID:g633870  
GENETICS  
#note gene name H4(D10S170)  
SUMMARY #length 585 #molecular-weight 65916 #checksum 4727

Query Match 70.5% Score 55; DB 13; Length 585;  
Best Local Similarity 61.5%; Pred. No. 1.44e+00;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Db 161 qhlegqefgvnk 173  
||| ||| |

QY 2 QHXSXQEQTFQLXK 14

RESULT 2

ENTRY #type complete  
TITLE E1 protein - human papillomavirus type 51  
ORGANISM #formal\_name human papillomavirus type 51  
#note host Homo sapiens (man)  
DATE 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 27-Jan-1995  
ACCESSIONS A40415  
REFERENCE A40415  
#authors Lungu, O.; Crum, C.P.; Silverstein, S.J.  
#journal J. Virol. (1991) 65:4216-4225  
#title Biologic properties and nucleotide sequence analysis of human papillomavirus type 51.  
#cross-references MUID:91303675

```

#accession      A40415
##status        translation not shown
##molecule_type DNA
##residues      1-634 #label LUN
##cross-references GB:M62877
CLASSIFICATION #superfamily papillomavirus E1 protein
KEYWORDS        early protein
SUMMARY         #length 634 #molecular-weight 71684 #checksum 6865

Query Match      67.9%; Score 53; DB 4; Length 634;
Best Local Similarity 54.5%; Pred. No. 3.38e+00;
Matches          6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 333 qhsfedstfel 343
|||::|||
QY 2 QHSXQEQTFQL 12

3
ULT
Y
LE
#accession      S36546
##status        #type complete
##molecule_type E1 protein - human papillomavirus type 26
##formal_name human papillomavirus type 26
##journal        20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
##accession      S36546
##residues      1-638 #label DEL
##cross-references EMBL:X74472
CLASSIFICATION #superfamily papillomavirus E1 protein
KEYWORDS        early protein
SUMMARY         #length 638 #molecular-weight 71956 #checksum 4174

Query Match      67.9%; Score 53; DB 8; Length 638;
Best Local Similarity 46.2%; Pred. No. 3.38e+00;
Matches          6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 337 ehsfddatfdisk 349
|||::|||
QY 2 QHSXQEQTFQLXK 14

4
ULT
Y
LE
#accession      A56158
##status        #type complete
##molecule_type eye development protein canoe - fruit fly (Drosophila melanogaster)
##formal_name Drosophila melanogaster
##journal        03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change
##accession      A56158
##residues      1-1893 #label MIY
##cross-references GB:D49534
CLASSIFICATION #superfamily papillomavirus E1 protein
KEYWORDS        early protein
SUMMARY         #length 1893 #molecular-weight 210186 #checksum 2528

Query Match      67.9%; Score 53; DB 12; Length 1893;
Best Local Similarity 63.6%; Pred. No. 3.38e+00;
Matches          7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```

```

Db 1179 qhslgggqfal 1189
|||::|||
QY 2 QHSXQEQTFQL 12

5
RESULT ENTRY
TITLE S43435 #type complete
nuclear receptor protein DR-78 - fruit fly (Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster
DATE 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
ACCESSIONS S43435
REFERENCE S43435
authors Martin-Blanco, E.; Kornberg, T.B.
#journal Biochim. Biophys. Acta (1993) 1216:339-341
#title DR-78, a novel Drosophila melanogaster genomic DNA fragment highly homologous to the DNA-binding domain of thyroid hormone-retinoic acid-vitamin D receptor subfamily.
#accession S43435
##status preliminary
##molecule_type DNA
##residues 1-113 #label MAR
##cross-references EMBL:X73045
GENETICS
#introns 68/3
SUMMARY #length 113 #molecular-weight 12523 #checksum 196

Query Match 66.7%; Score 52; DB 12; Length 113;
Best Local Similarity 54.5%; Pred. No. 5.14e+00;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 13 qhpgggsgfql 23
|||::|||
QY 2 QHSXQEQTFQL 12

6
RESULT ENTRY
TITLE S32723 #type fragment
DR-78 protein - fruit fly (Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster
DATE 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
ACCESSIONS S32723
REFERENCE S32723
authors Martin-Blanco, E.; Kornberg, T.B.
#submission submitted to the EMBL Data Library, April 1993
#description DR-78, a novel Drosophila melanogaster genomic DNA fragment highly homologous to retinoic acid receptors DNA binding domain.
#accession S32723
##status preliminary
##molecule_type DNA
##residues 1-113 #label MAR
##cross-references EMBL:X73045
GENETICS
#introns 68/3
SUMMARY #length 113 #checksum 9804

Query Match 66.7%; Score 52; DB 12; Length 113;
Best Local Similarity 54.5%; Pred. No. 5.14e+00;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 13 qhpgggsgfql 23
|||::|||
QY 2 QHSXQEQTFQL 12

7
RESULT ENTRY
TITLE A49070 #type complete
ecdysone-inducible protein E78A - fruit fly (Drosophila

```

```

#submission submitted to the EMBL Data Library, August 1993
#description Primer-directed sequencing of human papillomavirus types.
#accession S36505
#molecule_type DNA
#residues 1-631 ##label DEL
#cross-references EMBL:X7474
CLASSIFICATION #superfamily papillomavirus E1 protein
KEYWORDS early protein
SUMMARY #length 631 #molecular-weight 71809 #checksum 8022

Query Match 65.4%; Score 51; DB 8; Length 631;
Best Local Similarity 53.8%; Pred. No. 7.78e+00;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 329 qhsfcdqcfelsk 341
||| | | | |
QY 2 QHSXQEQTFQLXK 14

RESULT 10
ENTRY A49707 #type complete
TITLE phospholipase A2 receptor precursor - rabbit
ORGANISM rabbit
DATE 12-May-1994 #sequence_revision 12-May-1994 #text_change
09-Sep-1995
ACCESSIONS A49707
REFERENCE A49707
#authors Lambeau, G.; Ancian, P.; Barhanin, J.; Lazdunski, M.
#journal J. Biol. Chem. (1994) 269:1575-1578
#title Cloning and expression of a membrane receptor for secretory
phospholipases A-2.
#accession A49707
##status preliminary
##molecule_type mRNA
##residues 1-1458 ##label LAM
#cross-references GB:U03455
CLASSIFICATION #superfamily phospholipase A2 receptor; C-type lectin
homology; fibronectin type II repeat homology
KEYWORDS glycoprotein; receptor; skeletal muscle; tandem repeat;
transmembrane protein
FEATURE
176-217 #domain fibronectin type II repeat homology #label 2F1\
376-499 #domain C-type lectin homology #label LCH
SUMMARY #length 1458 #molecular-weight 167199 #checksum 6528

Query Match 64.1%; Score 50; DB 7; Length 1458;
Best Local Similarity 58.3%; Pred. No. 1.17e+01;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Db 849 hsaheqefilsk 860
||| | | | |
QY 3 HSAHEQEFQLXK 14

RESULT 11
ENTRY B33956 #type fragment
TITLE tyrocidine synthetase (EC 6.---) component 2 - Bacillus
brevis (fragment)
ORGANISM #formal_name Bacillus brevis
DATE 23-Mar-1990 #sequence_revision 23-Mar-1990 #text_change
21-Mar-1996
ACCESSIONS B33956
REFERENCE A33956
#authors Mittenhuber, G.; Weckermann, R.; Marahiel, M.A.
#journal J. Bacteriol. (1989) 171:4881-4887
#title Gene cluster containing the genes for tyrocidine synthetases
1 and 2 from Bacillus brevis: evidence for an operon.
#cross-references MUID:89359123
#accession B33956
##status preliminary; not compared with conceptual translation
##molecule_type DNA
##residues 1-245 ##label MIT

```

```

#submission submitted to the EMBL Data Library, August 1993
#description Primer-directed sequencing of human papillomavirus types.
#accession S36505
#molecule_type DNA
#residues 1-631 ##label DEL
#cross-references EMBL:X7474
CLASSIFICATION #superfamily papillomavirus E1 protein
KEYWORDS early protein
SUMMARY #length 631 #molecular-weight 71809 #checksum 8022

Query Match 65.4%; Score 51; DB 8; Length 631;
Best Local Similarity 53.8%; Pred. No. 7.78e+00;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 329 qhsfcdqcfelsk 341
||| | | | |
QY 2 QHSXQEQTFQLXK 14

RESULT 10
ENTRY A49707 #type complete
TITLE phospholipase A2 receptor precursor - rabbit
ORGANISM rabbit
DATE 12-May-1994 #sequence_revision 12-May-1994 #text_change
09-Sep-1995
ACCESSIONS A49707
REFERENCE A49707
#authors Lambeau, G.; Ancian, P.; Barhanin, J.; Lazdunski, M.
#journal J. Biol. Chem. (1994) 269:1575-1578
#title Cloning and expression of a membrane receptor for secretory
phospholipases A-2.
#accession A49707
##status preliminary
##molecule_type mRNA
##residues 1-1458 ##label LAM
#cross-references GB:U03455
CLASSIFICATION #superfamily phospholipase A2 receptor; C-type lectin
homology; fibronectin type II repeat homology
KEYWORDS glycoprotein; receptor; skeletal muscle; tandem repeat;
transmembrane protein
FEATURE
176-217 #domain fibronectin type II repeat homology #label 2F1\
376-499 #domain C-type lectin homology #label LCH
SUMMARY #length 1458 #molecular-weight 167199 #checksum 6528

Query Match 64.1%; Score 50; DB 7; Length 1458;
Best Local Similarity 58.3%; Pred. No. 1.17e+01;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Db 849 hsaheqefilsk 860
||| | | | |
QY 3 HSAHEQEFQLXK 14

RESULT 11
ENTRY B33956 #type fragment
TITLE tyrocidine synthetase (EC 6.---) component 2 - Bacillus
brevis (fragment)
ORGANISM #formal_name Bacillus brevis
DATE 23-Mar-1990 #sequence_revision 23-Mar-1990 #text_change
21-Mar-1996
ACCESSIONS B33956
REFERENCE A33956
#authors Mittenhuber, G.; Weckermann, R.; Marahiel, M.A.
#journal J. Bacteriol. (1989) 171:4881-4887
#title Gene cluster containing the genes for tyrocidine synthetases
1 and 2 from Bacillus brevis: evidence for an operon.
#cross-references MUID:89359123
#accession B33956
##status preliminary; not compared with conceptual translation
##molecule_type DNA
##residues 1-245 ##label MIT

```

```

CLASSIFICATION #superfamily alpha-aminoadipyl-cysteinyI-valine synthetase;
                acetate--CoA ligase homology; acyl carrier protein
KEYWORDS        homology; glucicidin S synthetase I repeat homology
SUMMARY         #length 245 #checksum 6032

Query Match     62.8%; Score 49; DB 6; Length 245;
Best Local Similarity 66.7%; Pred. No. 1.75e+01;
Matches         6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 117 gertfhak 125
      |||:|:|:|
QY 6 QEQTFQLXK 14

RESULT 12
ENTRY #type complete
TITLE methylgalactoside permease ATP-binding protein (mgla) homolog
ORGANISM - Haemophilus influenzae (strain Rd KW20)
DATE #formal_name Haemophilus influenzae
      18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change
      17-Nov-1995
ACCESSIONS H64096
REFERENCE A64000
#authors Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.;
          Kirkness, E.F.; Kervilange, A.R.; Bult, C.J.; Tomb, J.F.;
          Dougherty, B.A.; Merrick, J.M.; McKenney, K.; Sutton, G.;
          FitzHugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.;
          Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman,
          J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton,
          M.D.; Utterback, T.R.; Hanna, M.C.; Nguyen, D.T.; Saudek,
          D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann,
          J.L.; Geoghagen, N.S.M.; Gnehm, C.L.; McDonald, L.A.;
          Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.C.
#journal Science (1995) 269:496-512
#title Whole-genome random sequencing and assembly of Haemophilus
          influenzae Rd.
#accession H64096
#status preliminary
#molecule_type DNA
#residues 1-506 #label TIGR
#cross-references GB:L42023; TIGR:HI0823
#note named as homolog to a protein from Escherichia coli
CLASSIFICATION #superfamily malK protein homology
FEATURE        #domain malK protein homology #label MK1\
                #domain malK protein homology #label MK2
                #length 506 #molecular-weight 56567 #checksum 3128

Query Match     62.8%; Score 49; DB 9; Length 506;
Best Local Similarity 46.2%; Pred. No. 1.75e+01;
Matches         6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 276 qpsiqdvsefkr 288
      |||:|:|:|
QY 2 QHSXQEQTFQLXK 14

RESULT 13
ENTRY #type fragment
TITLE T-cell receptor delta chain precursor - sheep (fragment)
ORGANISM #formal_name Ovis orientalis aries, Ovis ammon aries
DATE #common_name domestic sheep
      03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change
      18-Nov-1994
ACCESSIONS S36325; S23040
REFERENCE S36287
#authors Hein, W.R.; Dudley, L.
#journal EMBO J. (1993) 12:715-724
#title Divergent evolution of T cell repertoires: extensive
          diversity and developmentally regulated expression of the
          sheep gamma-delta T cell receptor.
#accession S36325

```

```

#status preliminary; translation not shown
#molecule_type mRNA
#residues 1-139 #label HEI
#cross-references EMBL:Z12997
SUMMARY         #length 139 #checksum 4472

Query Match     61.5%; Score 48; DB 14; Length 139;
Best Local Similarity 54.5%; Pred. No. 2.61e+01;
Matches         6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 86 qhsrvsktfhl 96
      |||:|:|:|
QY 2 QHSXQEQTFQL 12

RESULT 14
ENTRY #type complete
TITLE involucrin - western tarsier
ORGANISM #formal_name Tarsius bancanus #common_name western tarsier
DATE 20-Feb-1993 #sequence_revision 23-Feb-1996 #text_change
      23-Aug-1996
ACCESSIONS A43704
REFERENCE A43704
#authors Djian, P.; Green, H.
          Proc. Natl. Acad. Sci. U.S.A. (1991) 88:5321-5325
#title involucrin gene of tarsoids and other primates: alternatives
          in evolution of the segment of repeats.
#accession A43704
#status translation not shown
#molecule_type DNA
#residues 1-387 #label DJI
#cross-references GB:M65124
COMMENT During the terminal differentiation of keratinocytes, this protein
          from the cytosol becomes cross-linked to membrane proteins by
          transglutaminase and incorporated into an insoluble, cross-linked
          envelope under the plasma membrane.
CLASSIFICATION #superfamily involucrin
KEYWORDS cornified cell envelope; duplication; epidermis; tandem
          repeat
SUMMARY         #length 387 #molecular-weight 45077 #checksum 6132

Query Match     61.5%; Score 48; DB 3; Length 387;
Best Local Similarity 46.2%; Pred. No. 2.61e+01;
Matches         6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

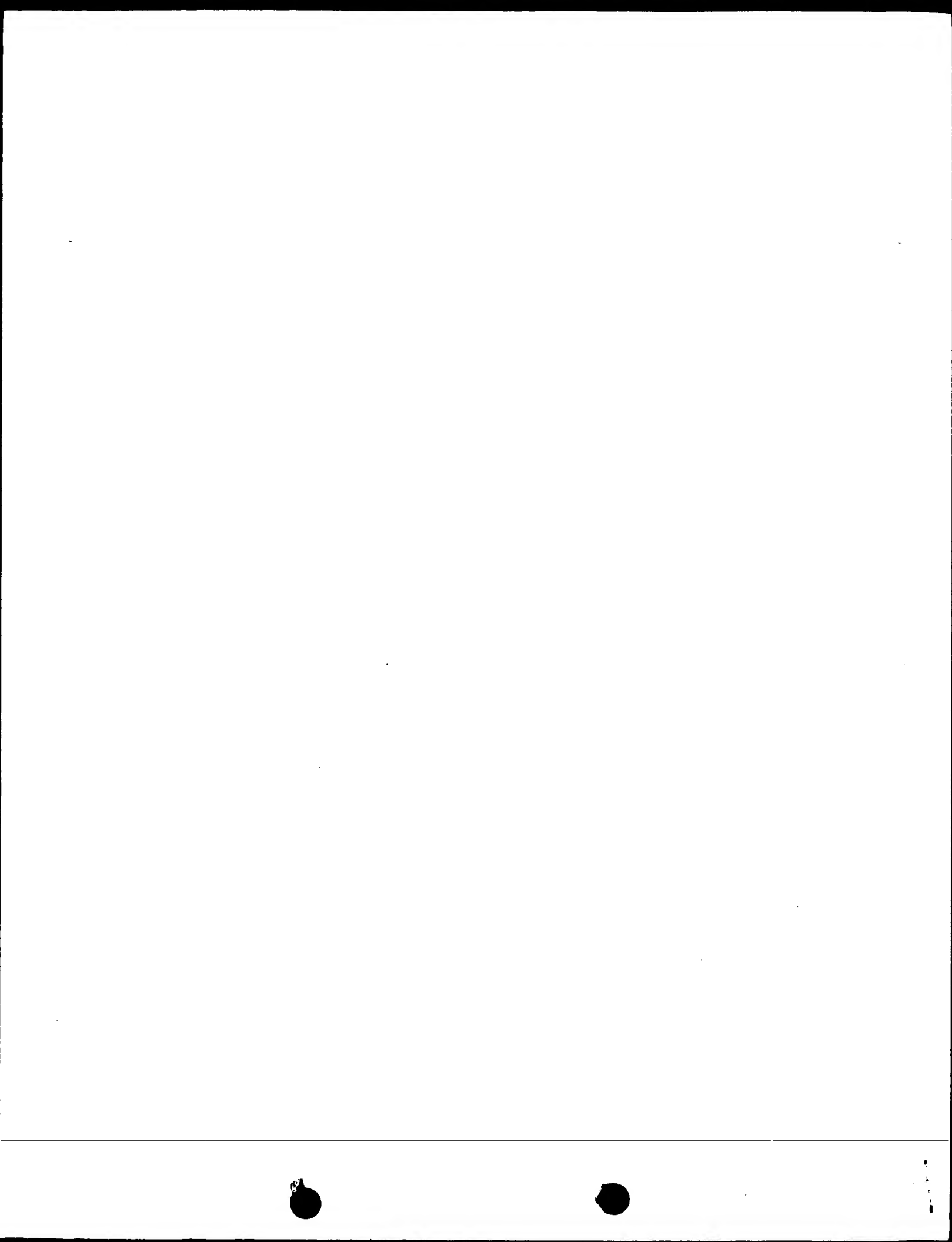
Db 235 gqaqgeqelhgk 247
      |||:|:|:|
QY 2 QHSXQEQTFQLXK 14

RESULT 15
ENTRY #type complete
TITLE involucrin - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Jul-1996 #sequence_revision 23-Aug-1996 #text_change
      23-Aug-1996
ACCESSIONS A49377
REFERENCE A49377
#authors Djian, P.; Phillips, M.; Easley, K.; Huang, E.; Simon, M.;
          Rice, R.H.; Green, H.
#journal Mol. Biol. Evol. (1993) 10:1136-1149
#title The involucrin genes of the mouse and the rat: study of their
          shared repeats.
#cross-references MUID:94104476
#accession A49377
#status translation not shown; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-467 #label RES
#cross-references GB:L28819; NID:g454418; CDS_PID:g454419
COMMENT During the terminal differentiation of keratinocytes, this protein
          from the cytosol becomes cross-linked to membrane proteins by
          transglutaminase and incorporated into an insoluble, cross-linked
          envelope under the plasma membrane.

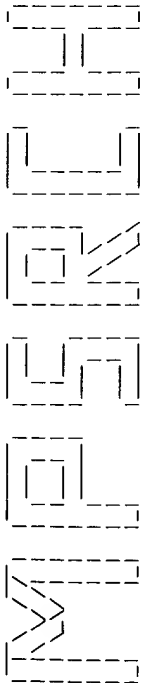
```

CLASSIFICATION #superfamily involucrin  
KEYWORDS cornified cell envelope; duplication; epidermis; tandem  
repeat  
SUMMARY #length 467 #molecular-weight 54919 #checksum 5202  
Query Match 61.5%; Score 48; DB 3; Length 467;  
Best Local Similarity 46.2%; Pred. No. 2.61e+01;  
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
Db 92 gqelqeqelhle 104  
Qy 2 QHSXOEQTFOLXK 14

Search completed: Tue Dec 2 14:44:51 1997  
Job time : 12 secs.



\*\*\*\*\*



Release 2.1d John F. Collins, Biocomputing Research Unit.  
Copyright (C) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by Intelligent, Inc.

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Tue Dec 2 14:44:12 1997; MasPar time 2.03 Seconds  
145.979 Million cell updates/sec  
Circular output not generated.

Title: >US-08-915-004-2  
Description: (1-14) from US08915004.pap  
Perfect Score: 78  
Sequence: 1 QHSHXQEQFQLXK 14

Scoring table: PAM 150  
Gap 15

Searched: 59021 seqs, 21210388 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot34  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11

Statistics: Mean 23.758; Variance 28.025; scale 0.848

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	53	67.9	634	10	VEI_HPV51	7.11e-01
2	53	67.9	634	10	VEI_HPV26	7.11e-01
3	52	66.7	864	3	E78A_DROME	1.17e+00
4	51	65.4	450	5	INVO_LEMCA	1.91e+00
5	51	65.4	631	10	VEI_HPV30	1.91e+00
6	50	64.1	1458	7	PA2R_RABIT	3.09e+00
7	49	62.8	506	6	MGIA_HAREIN	4.97e+00
8	48	61.5	387	5	INVO_TARBA	7.94e+00
9	48	61.5	467	5	INVO_MOUSE	7.94e+00
10	48	61.5	629	10	VEI_HPV31	7.94e+00
11	48	61.5	755	1	AMO_KLEAE	7.94e+00
12	48	61.5	757	1	AMO_ECOLI	7.94e+00
13	48	61.5	786	3	EXOP_RHIME	7.94e+00
14	47	60.3	335	9	TRPD_BUCAP	1.26e+01
15	47	60.3	337	5	KAPC_ASCSU	1.26e+01
16	47	60.3	426	9	TEF1_MOUSE	1.26e+01
17	47	60.3	426	9	TEF1_HUMAN	1.26e+01
18	47	60.3	570	11	YNE3_CAEEL	1.26e+01
19	47	60.3	923	5	HXK3_HUMAN	1.26e+01
20	46	59.0	154	1	AROD_ACTPL	1.98e+01
21	46	59.0	189	5	INAI_FIG	1.98e+01
22	46	59.0	275	4	FRAL_RAT	1.98e+01

23	46	59.0	494	2	CPA6_HUMAN	CYTOCHROME P450 IIA6	1.98e+01
24	46	59.0	530	7	PPB2_HUMAN	ALKALINE PHOSPHATASE,	1.98e+01
25	46	59.0	532	7	PPB2_HUMAN	ALKALINE PHOSPHATASE,	1.98e+01
26	46	59.0	535	7	PPB3_HUMAN	ALKALINE PHOSPHATASE,	1.98e+01
27	46	59.0	535	7	PPB1_HUMAN	ALKALINE PHOSPHATASE,	1.98e+01
28	46	59.0	644	10	VEI_HPV58	REPLICATION PROTEIN E	1.98e+01
29	46	59.0	713	7	NUB2_YEAST	NUCLEOPORIN NUP82 (NU	1.98e+01
30	46	59.0	747	1	AMD1_HUMAN	AMP DEAMINASE 1 (EC 3	1.98e+01
31	46	59.0	750	8	PSM_HUMAN	PROSTATE-SPECIFIC MEM	1.98e+01
32	46	59.0	1248	9	TOPG_SULAC	REVERSE GYRASE (CONTA	1.98e+01
33	46	59.0	1511	7	PDR5_YEAST	SUPPRESSOR OF TOXICIT	1.98e+01
34	45	57.7	118	11	YYAQ_BACSU	HYPOTHETICAL 13.9 KD	3.09e+01
35	45	57.7	190	3	ESS1_YEAST	ESS1 PROTEIN (PROCESS	3.09e+01
36	45	57.7	254	4	GLCC_ECOLI	GLC OPERON TRANSCRIPT	3.09e+01
37	45	57.7	372	5	HMEN_BOMMO	SEGMENTATION POLARITY	3.09e+01
38	45	57.7	438	9	TEFI_CHICK	TRANSCRIPTIONAL ENHAN	3.09e+01
39	45	57.7	515	2	CC4H_HUMAN	CDC4-LIKE PROTEIN (FR	3.09e+01
40	45	57.7	559	10	UROT_RAT	TISSUE PLASMINOGEN AC	3.09e+01
41	45	57.7	559	10	UROT_MOUSE	TISSUE PLASMINOGEN AC	3.09e+01
42	45	57.7	771	11	YJC0_YEAST	HYPOTHETICAL 86.1 KD	3.09e+01
43	45	57.7	1059	7	POL2_DROME	RETROVIRUS-RELATED PO	3.09e+01
44	45	57.7	1394	3	E75B_DROME	ECDYSONE-INDUCIBLE PR	3.09e+01
45	45	57.7	2314	8	PTP2_HUMAN	PROTEIN-TYROSINE PHOS	3.09e+01

ALIGNMENTS

RESULT 1	STANDARD;	PRT;	634 AA.
ID VEI_HPV51			
AC P26544;			
DT 01-AUG-1992 (REL. 23, CREATED)			
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)			
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)			
DE REPLICATION PROTEIN EL.			
GN EL.			
OS HUMAN PAPILLOMAVIRUS TYPE 51.			
OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PAPAPOVIRIDAE; PAPILLOMAVIRUSES.			
RN [1]			
RP SEQUENCE FROM N.A.			
RX MEDLINE; 91303675.			
RA LUNGU O., CRUM C.P., SILVERSTEIN S.J.;			
RL J. VIROL. 65:4216-4225(1991).			
CC -!- FUNCTION: ATP-DEPENDENT DNA HELICASE REQUIRED FOR DNA REPLICATION.			
CC -!- SUBCELLULAR LOCATION: NUCLEAR.			
DR EMBL; M62877; -; NOT_ANNOTATED_CDS.			
DR PIR; A40415; W1WLS1.			
KW EARLY PROTEIN; DNA REPLICATION; HELICASE; ATP-BINDING;			
KW NUCLEAR PROTEIN.			
FT NP_BIND 463 470 ATP (POTENTIAL).			
SQ SEQUENCE 634 AA; 71684 MW; DCCE91ED CRC32;			

Query Match 67.9%; Score 53; DB 10; Length 634;  
Best Local Similarity 54.5%; Pred. No. 7.11e-01;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 333 qhsfedstfel 343

QY 2 QHSHXQEQTFOL 12

RESULT 2	STANDARD;	PRT;	638 AA.
ID VEI_HPV26			
AC P36722;			
DT 01-JUN-1994 (REL. 29, CREATED)			
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)			
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)			
DE REPLICATION PROTEIN EL.			
GN EL.			
OS HUMAN PAPILLOMAVIRUS TYPE 26.			
OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PAPAPOVIRIDAE; PAPILLOMAVIRUSES.			
RN [1]			
RP SEQUENCE FROM N.A.			
RX MEDLINE; 94265501.			

RA DELIUS H., HOFMANN B.;  
 RL CURR. TOP. MICROBIOL. IMMUNOL. 186:13-31(1994).  
 CC -1- FUNCTION: ATP-DEPENDENT DNA HELICASE REQUIRED FOR DNA REPLICATION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL; X74472; G396959; -.  
 DR PIR; J36346; S36546.  
 KW EARLY PROTEIN; DNA REPLICATION; HELICASE; ATP-BINDING;  
 FT NP\_BIND 467 474 ATP (POTENTIAL).  
 SQ SEQUENCE 638 AA; 71956 MW; 5E36534F CRC32;

Query Match 67.9%; Score 53; DB 10; Length 638;  
 Best Local Similarity 46.2%; Pred. No. 7.11e-01;  
 Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 337 ehsfddatfdlslk 349  
 2 QHSXQEQTFQLXK 14

RESULT 3  
 ID E78A.DROME STANDARD; PRT; 864 AA.  
 AC P45447;  
 DT 01-NOV-1995 (REL. 32, CREATED)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE ECDYSONE-INDUCIBLE PROTEIN E78-A (DR-78).  
 GN EIP78C OR E78A.  
 OS DROSOPHILA MELANOGASTER (FRUIT FLY).  
 OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CANTON-S;  
 RX MEDLINE; 94006562.  
 RA STONE B.L., THUMMEL C.S.;  
 RL CELL 75:307-320(1993).  
 RN [2]  
 RP SEQUENCE OF 321-433 FROM N.A.  
 RX MEDLINE; 94060116.  
 RA MARTIN-BLANCO E., KORNBERG T.B.;  
 RL BIOCHIM. BIOPHYS. ACTA 1216:339-341(1993).  
 CC -1- FUNCTION: INDUCES THE EARLY LATE PUFF 78C WHICH TRIGGERS PUPARIUM FORMATION AND DEVELOPMENT.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).  
 CC -1- DEVELOPMENTAL STAGE: THE LONGER FORM, E78A, IS EXPRESSED ONLY IN MID-PUPAL STAGES, WHILE THE SHORTER FORM, E78B, IS MAXIMALLY EXPRESSED IN NEWLY FORMED PREPUPAE.  
 CC -1- INDUCTION: BOTH FORMS REQUIRE ECDYSONE FOR ACTIVITY. E78B ALSO REQUIRES ECDYSONE-INDUCED PROTEINS FOR MAXIMAL EXPRESSION.  
 CC -1- ALTERNATIVE PRODUCTS: THE DIFFERENT FORMS OF PROTEIN E78 ARE PROBABLY PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.  
 CC -1- SIMILARITY: TO OTHER MEMBERS OF THE STEROID/THYROID/RETINOIC NUCLEAR HORMONE RECEPTORS.  
 DR EMBL; U01087; G514833; -.  
 DR EMBL; U01088; G514834; -.  
 DR EMBL; X73045; E74981; -.  
 DR FLYBASE; FBGN004865; EIP78C.  
 DR PROSITE; PS00031; STEROID\_FINGER.  
 KW TRANSCRIPTION REGULATION; DNA-BINDING; NUCLEAR PROTEIN;  
 KZ ZINC-FINGER.  
 FT DNA\_BIND 367 432 C4-TYPE ZINC FINGERS (TWO).  
 FT ZN\_FING 367 387  
 FT ZN\_FING 403 427 C4-TYPE.  
 FT ZN\_FING 403 427 C4-TYPE.  
 FT DOMAIN 182 188 POLY-GLU.  
 FT DOMAIN 192 202 POLY-GLN.  
 FT DOMAIN 240 247 POLY-SER.  
 FT DOMAIN 271 279 POLY-SER.  
 FT DOMAIN 312 315 POLY-GLN.  
 FT DOMAIN 321 333 POLY-GLN.  
 FT DOMAIN 336 339 POLY-GLN.  
 FT DOMAIN 346 349 POLY-SER.  
 FT DOMAIN 354 357 POLY-ASN.

FT DOMAIN 481 486 POLY-GLN.  
 FT DOMAIN 490 500 POLY-GLN.  
 FT DOMAIN 546 554 POLY-ASN.  
 FT VARSPLIC 1 474 MISSING (IN TRUNCATED FORM E78B).  
 FT CONFLICT 321 331 QLOQQQHQQQ -> SCNSSSTSSR (IN REF. 2).  
 FT CONFLICT 430 433 AGMS -> VGMM (IN REF. 2).  
 SQ SEQUENCE 864 AA; 95865 MW; 5EEB72CO CRC32;

Query Match 66.7%; Score 52; DB 3; Length 864;  
 Best Local Similarity 54.5%; Pred. No. 1.17e+00;  
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 333 qhpgqgqsgf1 343  
 2 QHSXQEQTFQL 12

RESULT 4  
 ID INVO.LEMCA STANDARD; PRT; 450 AA.  
 AC P14590;  
 DT 01-JAN-1990 (REL. 13, CREATED)  
 DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
 DE INVOLUCRIN.  
 GN IVL.  
 OS LEMUR CATTA (RING-TAILED LEMUR).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 88295123.  
 RA TSENG H., GREEN H.;  
 RL CELL 54:491-496(1988).  
 CC -1- FUNCTION: INVOLUCRIN IS A KERATINOCYTE PROTEIN THAT FIRST APPEARS IN THE CELL CYTOSOL, BUT ULTIMATELY BECOMES CROSS-LINKED TO MEMBRANE PROTEINS BY TRANSGLUTAMINASE. ALL THAT RESULTS IN THE FORMATION OF AN INSOLUBLE ENVELOPE BENEATH THE PLASMA MEMBRANE.  
 CC -1- TISSUE SPECIFICITY: PRESENT IN KERATINOCYTES OF EPIDERMIS AND OTHER STRATIFIED SQUAMOUS EPITHELIA.  
 DR EMBL; M21864; G340630; -.  
 DR PIR; A43733; A43733.  
 DR PROSITE; PS00795; INVOLUCRIN.  
 KW KERATINOCYTE; REPEAT.  
 SQ SEQUENCE 450 AA; 50445 MW; BB180C73 CRC32;

Query Match 65.4%; Score 51; DB 5; Length 450;  
 Best Local Similarity 46.2%; Pred. No. 1.91e+00;  
 Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 151 qqapqeqeqlhlgk 163  
 2 QHSXQEQTFQLXK 14

RESULT 5  
 ID VEL\_HPV30 STANDARD; PRT; 631 AA.  
 AC Q05112;  
 DT 01-FEB-1994 (REL. 28, CREATED)  
 DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE REPLICATION PROTEIN E1.  
 GN E1.  
 OS HUMAN PAPILLOMAVIRUS TYPE 30.  
 OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PAPAPOVIRIDAE; PAPILLOMAVIRUSES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 94265501.  
 RA DELIUS H., HOFMANN B.;  
 RL CURR. TOP. MICROBIOL. IMMUNOL. 186:13-31(1994).  
 RN [2]  
 RP SEQUENCE OF 358-409 FROM N.A.  
 RX MEDLINE; 92407963.  
 RA CHAN S.Y., BERNARD H.U., ONG C.K., CHAN S.P., BIRGIT H., DELIUS H.;



RL J. VIROL. 66:5714-5725(1992).  
 CC -!- FUNCTION: ATP-DEPENDENT DNA HELICASE REQUIRED FOR DNA REPLICATION.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL: X74474; G396976; -.  
 DR EMBL: M96304; G333093; -.  
 DR PIR: S36505; S36505.  
 KW EARLY PROTEIN; DNA REPLICATION; HELICASE; ATP-BINDING;  
 FT NUCLEAR PROTEIN.  
 FT NP\_BIND 459 466 ATP (POTENTIAL).  
 SQ SEQUENCE 631 AA; 71809 MW; 753395D7 CRC32;

Query Match 65.4%; Score 51; DB 10; Length 631;  
 Best Local Similarity 53.8%; Pred. No. 1.91e+00;  
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 329 qhsfqdcqfslk 341  
 ||| | : | : |  
 QY 2 QHSXQEQTFOLXK 14

RESULT 6  
 ID PA2R\_RABIT STANDARD; PRT: 1458 AA.  
 AC P49260;  
 DT 01-FEB-1996 (REL. 33, CREATED)  
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE 180 KD SECRETORY PHOSPHOLIPASE A2 RECEPTOR PRECURSOR (PLA2-R).  
 OS ORYCTOLAGUS CUNICULUS (RABBIT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; LAGOMORPHA.  
 RN [1]  
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.  
 RC TISSUE-SKELETAL MUSCLE;  
 RX MEDLINE; 94124484.  
 RA LAMBEAU G., ANCIAN P., BARHANIN J., LAZDUNSKI M.;  
 J. BIOL. CHEM. 269:1575-1578(1994).  
 CC -!- FUNCTION: MAY HAVE A KEY ROLE IN NORMAL AND PATHOLOGICAL ACTIONS  
 CC OF SECRETORY PHOSPHOLIPASE A2. ALSO BINDS TO SNAKE PA2-LIKE  
 CC TOXINS.  
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -!- TISSUE SPECIFICITY: LONG, SKELETAL MUSCLE, BRAIN, KIDNEY AND  
 CC HEART.  
 CC -!- SIMILARITY: CONTAINS 8 C-TYPE LECTIN FAMILY DOMAINS.  
 DR EMBL: U03455; G456376; -.  
 KW SIGNAL; RECEPTOR; TRANSMEMBRANE; REPEAT; GLYCOPROTEIN; LECTIN.  
 FT SIGNAL 1 23 POTENTIAL.  
 FT CHAIN 24 1458 180 KD SECRETORY PHOSPHOLIPASE A2  
 RECEPTOR.  
 FT DOMAIN 24 1393 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 1394 1416 POTENTIAL.  
 FT DOMAIN 1417 1458 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 24 165 CYS-RICH.  
 FT DOMAIN 165 220 FIBRONECTIN TYPE-II.  
 FT DOMAIN 227 356 C-TYPE LECTIN 1 (LONG FORM).  
 FT DOMAIN 374 502 C-TYPE LECTIN 2 (LONG FORM).  
 FT DOMAIN 511 645 C-TYPE LECTIN 3 (LONG FORM).  
 FT DOMAIN 660 798 C-TYPE LECTIN 4 (LONG FORM).  
 FT DOMAIN 815 939 C-TYPE LECTIN 5 (LONG FORM).  
 FT DOMAIN 954 1098 C-TYPE LECTIN 6 (LONG FORM).  
 FT DOMAIN 1117 1231 C-TYPE LECTIN 7 (LONG FORM).  
 FT DOMAIN 1243 1376 C-TYPE LECTIN 8 (LONG FORM).  
 FT CARBOHYD 91 91 POTENTIAL.  
 FT CARBOHYD 408 408 POTENTIAL.  
 FT CARBOHYD 431 431 POTENTIAL.  
 FT CARBOHYD 452 452 POTENTIAL.  
 FT CARBOHYD 471 471 POTENTIAL.  
 FT CARBOHYD 582 582 POTENTIAL.  
 FT CARBOHYD 725 725 POTENTIAL.  
 FT CARBOHYD 778 778 POTENTIAL.  
 FT CARBOHYD 907 907 POTENTIAL.  
 FT CARBOHYD 981 981 POTENTIAL.  
 FT CARBOHYD 1054 1054 POTENTIAL.  
 FT CARBOHYD 1106 1106 POTENTIAL.

FT CARBOHYD 1121 1121 POTENTIAL.  
 FT CARBOHYD 1130 1130 POTENTIAL.  
 FT CARBOHYD 1319 1319 POTENTIAL.  
 FT CARBOHYD 1439 1439 POTENTIAL.  
 SQ SEQUENCE 1458 AA; 167199 MW; FA57C2CA CRC32;  
 Query Match 64.1%; Score 50; DB 7; Length 1458;  
 Best Local Similarity 58.3%; Pred. No. 3.09e+00;  
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Db 849 hsaheqefilsk 860  
 || | : | : |  
 QY 3 HXSQEQTFOLXK 14

RESULT 7  
 ID MGLA\_HAEIN STANDARD; PRT: 506 AA.  
 AC P44884;  
 DT 01-NOV-1995 (REL. 32, CREATED)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 DE GALACTOSIDE TRANSPORT ATP-BINDING PROTEIN MGLA.  
 GN MGLA OR H10823.  
 OS HAEMOPHILUS INFLUENZAE.  
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
 OC PASTEURILLACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-RD / KW20;  
 RX MEDLINE; 95350630.  
 RA FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,  
 RA FLEISCHMANN R.D., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,  
 RA MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,  
 RA SCOTT J.D., SHIRLEY R., LIU L.-I., GLODEK A., KELLEY J.M.,  
 RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOM E., COTTON M.D.,  
 RA UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,  
 RA FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,  
 RA GNEM C.L., McDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,  
 RA VENTER J.C.;  
 RL SCIENCE 269:496-512(1995).  
 CC -!- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM  
 CC FOR GALACTOSIDES; PROBABLY RESPONSIBLE FOR ENERGY COUPLING TO  
 CC THE TRANSPORT SYSTEM (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: INNER MEMBRANE-ASSOCIATED (POTENTIAL).  
 CC -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY  
 CC (ABC TRANSPORTERS).  
 DR EMBL: L45461; G1005849; -.  
 DR EMBL: U32764; G925867; -.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER.  
 KW TRANSPORT; SUGAR TRANSPORT; INNER MEMBRANE; ATP-BINDING.  
 FT NP\_BIND 46 53 ATP (BY SIMILARITY).  
 SQ SEQUENCE 506 AA; 56567 MW; 8964ECA2 CRC32;

Query Match 62.8%; Score 49; DB 6; Length 506;  
 Best Local Similarity 46.2%; Pred. No. 4.97e+00;  
 Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 276 qpsiqvsvfslk 288  
 || | : | : |  
 QY 2 QHSXQEQTFOLXK 14

RESULT 8  
 ID INVO\_TARBA STANDARD; PRT: 387 AA.  
 AC P24711;  
 DT 01-MAR-1992 (REL. 21, CREATED)  
 DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
 DE INVOLUCRIN.  
 GN IVL.  
 OS TARSUS BANCANUS (WESTERN TARSIER).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.

RN SEQUENCE FROM N.A.  
 RC TISSUE=LIVER;  
 RX MEDLINE; 91271381.  
 RA DJIAN P., GREEN H.;  
 RL PROC NATL ACAD SCI U.S.A. 88:5321-5325(1991).  
 CC -!- FUNCTION: INVOLUCRIN IS A KERATINOCYTE PROTEIN THAT FIRST APPEARS  
 CC IN THE CELL CYTOSOL, BUT ULTIMATELY BECOMES CROSS-LINKED TO  
 CC MEMBRANE PROTEINS BY TRANSGLUTAMINASE. ALL THAT RESULTS IN THE  
 CC FORMATION OF AN INSOLUBLE ENVELOPE BENEATH THE PLASMA MEMBRANE.  
 CC -!- TISSUE SPECIFICITY: PRESENT IN KERATINOCYTES OF EPIDERMIS AND  
 CC OTHER STRATIFIED SQUAMOUS EPITHELIA.  
 DR EMBL; M65124; G343452; -.  
 DR PIR; A43704; A43704.  
 DR PROSITE; PS00795; INVOLUCRIN.  
 KW KERATINOCYTE; REPEAT.  
 SQ SEQUENCE 387 AA; 45077 MW; 2800B15B CRC32;  
 Query Match 61.5%; Score 48; DB 5; Length 387;  
 Best Local Similarity 46.2%; Pred. No. 7.94e+00;  
 Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
 Db 235 ggeaqeqlhlgk 247  
 QY 2 QHSXQEQTFQLXK 14  
 RESULT 9  
 ID INVO\_MOUSE STANDARD; PRT; 467 AA.  
 AC P48997;  
 DT 01-FEB-1996 (REL. 33, CREATED)  
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
 DE INVOLUCRIN.  
 DE INVOLUCRIN.  
 GN IVL.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NH SWISS;  
 RX MEDLINE; 94104476.  
 RA DJIAN P., PHILLIPS M., EASLEY K., HUANG E., SIMON M., RICE R.H.,  
 RA GREEN H.;  
 RL MOL BIOL EVOL 10:1136-1149(1993).  
 CC -!- FUNCTION: INVOLUCRIN IS A KERATINOCYTE PROTEIN THAT FIRST APPEARS  
 CC IN THE CELL CYTOSOL, BUT ULTIMATELY BECOMES CROSS-LINKED TO  
 CC MEMBRANE PROTEINS BY TRANSGLUTAMINASE. ALL THAT RESULTS IN THE  
 CC FORMATION OF AN INSOLUBLE ENVELOPE BENEATH THE PLASMA MEMBRANE.  
 CC -!- TISSUE SPECIFICITY: PRESENT IN KERATINOCYTES OF EPIDERMIS AND  
 CC OTHER STRATIFIED SQUAMOUS EPITHELIA.  
 DR EMBL; L28819; G454419; -.  
 DR KERATINOCYTE; REPEAT.  
 SQ SEQUENCE 467 AA; 54919 MW; 39DC0AFE CRC32;  
 Query Match 61.5%; Score 48; DB 5; Length 467;  
 Best Local Similarity 46.2%; Pred. No. 7.94e+00;  
 Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
 Db 92 qqelqeeqlhlek 104  
 QY 2 QHSXQEQTFQLXK 14  
 RESULT 10  
 ID VEL\_HPV31 STANDARD; PRT; 629 AA.  
 AC P17382;  
 DT 01-AUG-1990 (REL. 15, CREATED)  
 DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE REPLICATION PROTEIN EL.  
 GN EL.  
 OS HUMAN PAPILLOMAVIRUS TYPE 31.

OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PAPAPOVIRIDAE; PAPILLOMAVIRUSES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 89299478.  
 RA GOLDSBOROUGH M.D., DISILVESTRE D., TEMPLE G.F., LORINCZ A.T.;  
 RL VIROLOGY 171:306-311(1989).  
 CC -!- FUNCTION: ATP-DEPENDENT DNA HELICASE REQUIRED FOR DNA REPLICATION.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL; J04353; G459918; -.  
 DR PIR; C32444; W1WL31.  
 KW EARLY PROTEIN; DNA REPLICATION; HELICASE; ATP-BINDING;  
 KW NUCLEAR PROTEIN.  
 FT NP\_BIND 457 464 ATP (POTENTIAL).  
 SQ SEQUENCE 629 AA; 71227 MW; DA96E287 CRC32;  
 Query Match 61.5%; Score 48; DB 10; Length 629;  
 Best Local Similarity 54.5%; Pred. No. 7.94e+00;  
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 Db 327 qhsfndttfdl 337  
 QY 2 QHSXQEQTFQL 12  
 RESULT 11  
 ID AMO\_KLEAE STANDARD; PRT; 755 AA.  
 AC P49250;  
 DT 01-FEB-1996 (REL. 33, CREATED)  
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
 DE AMINE OXIDASE PRECURSOR (EC 1.4.3.6) (MONAMINE OXIDASE) (TYRAMINE  
 DE OXIDASE).  
 GN MAOA OR TYNA.  
 OS KLEBSIELLA AEROGES.  
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
 OC ENTEROBACTERIACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-41.  
 RC STRAIN=W70;  
 RX MEDLINE; 92210491.  
 RA SUGINO H., SASAKI M., AZAKAMI H., YAMASHITA M., MUROOKA Y.;  
 RL J. BACTERIOL. 174:2485-2492(1992).  
 CC -!- FUNCTION: ACTIVE ON TYRAMINE, TRYPTAMINE, BETA-PHENETHYLAMINE AND  
 CC DOPAMINE.  
 CC -!- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) = RCHO + NH(3) +  
 CC H(2)O(2).  
 CC -!- COFACTOR: COPPER AND TOPAQUINONE.  
 CC -!- SUBCELLULAR LOCATION: PERIPLASMIC.  
 CC -!- INDUCTION: BY TYRAMINE.  
 CC -!- SIMILARITY: TO OTHER COPPER/TOPAQUINONE OXIDASES.  
 DR EMBL; D10208; G216723; -.  
 DR OXIDOREDUCTASE; COPPER; TPQ; PERIPLASMIC; SIGNAL.  
 FT SIGNAL 1 30  
 FT CHAIN 31 755 COPPER AMINE OXIDASE.  
 FT ACT\_SITE 413 413 CATALYTIC BASE (BY SIMILARITY).  
 FT BINDING 496 496 TOPAQUINONE (BY SIMILARITY).  
 FT METAL 554 554 COPPER (POTENTIAL).  
 FT METAL 556 556 COPPER (POTENTIAL).  
 FT METAL 719 719 COPPER (POTENTIAL).  
 SQ SEQUENCE 755 AA; 83576 MW; 735E26ED CRC32;  
 Query Match 61.5%; Score 48; DB 1; Length 755;  
 Best Local Similarity 50.0%; Pred. No. 7.94e+00;  
 Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 Db 110 qsgldgtfgev 121  
 QY 3 HSXQEQTFQLXK 14  
 RESULT 12  
 ID AMO\_ECOLI STANDARD; PRT; 757 AA.  
 AC P46883;

```

RP SEQUENCE FROM N.A.
RC STRAIN-1021;
RX MEDLINE; 94042870.
RA GLUCKSMANN M.A., REUBER T.L., WALKER G.C.;
RL J. BACTERIOL. 175:7045-7055(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-2011;
RX MEDLINE; 94067019.
RA BECKER A., KLEICKMANN A., KELLER M., ARNOLD W., PUEHLER A.;
RL MOL. GEN. GENET. 241:367-379(1993).
RC CC -1- PATHWAY: EXOPOLYSACCHARIDE BIOSYNTHESIS.
CC CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
DR ENBL; L20758; G393241; -
DR ENBL; Z22636; G605655; -
DR PIR; B43349; B49349.
DR PIR; S37031; S37031.
DR PIR; S39960; S39960.
DR KW EXOPOLYSACCHARIDE SYNTHESIS; TRANSPORT; TRANSMEMBRANE.
FT TRANSMEM 43 66
FT TRANSNEM 690 711
FT SEQUENCE 786 AA; 86123 MW; A45B9144 CRC32;
SQ SEQUENCE 786 AA; 86123 MW; A45B9144 CRC32;

Query Match 61.5%; Score 48; DB 3; Length 786;
Best Local Similarity 30.8%; Pred. No. 7.94e+00;
Matches 4; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Db 429 eqatgggsfplak 441
Qy 2 QHSXOEQTFQLXK 14

RESULT 14
ID TRPD.BUCAP STANDARD; PRT; 335 AA.
AC P42392;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE ANTHRANILATE PHOSPHORIBOSYLTRANSFERASE (EC 2.4.2.18).
GN TRPD.
OS BUCHNERA APHIDICOLA.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; UNCERTAIN.
RN [1]
RP SEQUENCE FROM N.A.
RA MUMSON M.A., BAUMANN P.;
RL SUBMITTED (DEC-1992) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- CATALYTIC ACTIVITY: ANTHRANILATE + PHOSPHORIBOSYL-PYROPHOSPHATE =
CC N-5'-PHOSPHORIBOSYL-ANTHRANILATE + PYROPHOSPHATE.
CC -1- PATHWAY: SECOND STEP IN BIOSYNTHESIS OF TRYPTOPHAN.
DR ENBL; Z19055; G396656; -
DR TRYPTOPHAN BIOSYNTHESIS; TRANSFERASE; GLYCOSYLTRANSFERASE.
SQ SEQUENCE 335 AA; 38090 MW; 74288035 CRC32;

Query Match 60.3%; Score 47; DB 9; Length 335;
Best Local Similarity 55.6%; Pred. No. 1.26e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 16 qeesyqlfx 24
Qy 6 QEQTFLXK 14

RESULT 15
ID KAPC.ASCSU STANDARD; PRT; 337 AA.
AC P49673;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE CAMP-DEPENDENT PROTEIN KINASE CATALYTIC SUBUNIT (EC 2.7.1.37) (PKA C).
GN CAPK.
OS ASCARIS SUUM (PIG ROUNDWORM) (ASCARIS LUMBRICOIDES).
OS EUKARYOTA; METAZOA; ACCELOMATES; NEMATODA; SECERNENTEA; ASCARIDIDA.
RN [1]

```

Search completed: Tue Dec 2 14:44:21 1997  
Job time : 9 secs.

\*\*\*\*\*  
M P E R L H  
\*\*\*\*\*  
(TM)

Release 2.1d John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Tue Dec 2 14:46:34 1997; MasPar time 2.02 Seconds  
73.024 Million cell updates/sec  
ular output not generated.

Title: >US-08-915-004-3  
Description: (1-12) from US08915004.pep  
Perfect Score: 94  
Sequence: 1 XIRFLHSFTMYK 12

Scoring table: PAM 150  
Gap 15

Searched: 101610 seqs, 12294212 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq28  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21

Statistics: Mean 18.812; Variance 62.835; scale 0.299

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

ult No.	Score	Query Match	Length	ID	Description	Pred. No.
1	94	100.0	12	20	R99923 Osteoclastogenesis in	5.77e-03
2	94	100.0	326	20	R99940 Mutated OCIF, OCIF-DD	5.77e-03
3	94	100.0	327	20	R99941 Mutated OCIF, OCIF-DD	5.77e-03
4	94	100.0	359	20	R99939 Mutated OCIF, OCIF-DC	5.77e-03
5	94	100.0	359	20	R99937 Mutated OCIF, OCIF-DC	5.77e-03
6	94	100.0	360	20	R99936 Mutated OCIF, OCIF-DC	5.77e-03
7	94	100.0	360	20	R99938 Mutated OCIF, OCIF-DC	5.77e-03
8	94	100.0	380	20	R99924 Mature osteoclastogen	5.77e-03
9	94	100.0	393	20	R99942 Mutated OCIF, OCIF-CL	5.77e-03
10	94	100.0	399	20	R99948 Mutated OCIF, OCIF-CL	5.77e-03
11	94	100.0	401	20	R99934 Mutated OCIF, OCIF-C2	5.77e-03
12	94	100.0	401	20	R99932 Mutated OCIF, OCIF-C2	5.77e-03
13	94	100.0	401	20	R99935 Mutated OCIF, OCIF-C2	5.77e-03
14	94	100.0	401	20	R99931 Mutated OCIF, OCIF-C1	5.77e-03
15	94	100.0	401	20	R99925 Full length osteoclas	5.77e-03
16	92	97.9	401	20	R99933 Mutated OCIF, OCIF-C2	9.54e-03
17	60	63.8	1189	11	R56496 TATA-binding protein-	2.11e+01
18	56	59.6	1165	1	R03420 bt4 Gene encoding Bt4	5.18e+01
19	56	59.6	1165	2	R10129 Lepidopteran-active t	5.18e+01
20	56	59.6	1165	2	R08200 bt4 Prototoxin gene pro	5.18e+01

21	56	59.6	1165	2	R10192 Insecticidal crystal	5.18e+01
22	56	59.6	1165	13	R72481 Bt4 protoxin.	5.18e+01
23	56	59.6	1165	8	R39756 Delta endotoxin.	5.18e+01
24	54	57.4	155	7	R34720 Bacillus subtilis srf	8.06e+01
25	54	57.4	325	14	R79502 Human melanocortin re	8.06e+01
26	54	57.4	325	9	R49726 Sequence of a polypep	8.06e+01
27	52	55.3	578	2	R10098 Alpha subunit of Alp-	1.25e+02
28	51	54.3	335	3	R12222 Conglutinin.	1.55e+02
29	51	54.3	371	15	R75642 Bovine conglutinin.	1.55e+02
30	51	54.3	882	8	R43959 N-heparan sulphate su	1.55e+02
31	50	53.2	331	2	R70328 Biotin-synthase gene	1.92e+02
32	50	53.2	337	1	R81191 Biotin synthetase (Bi	1.92e+02
33	49	52.1	223	3	R12899 Oncostatin M deletion	2.37e+02
34	49	52.1	227	3	R12897 Oncostatin M substit	2.37e+02
35	49	52.1	252	17	R88202 Human oncostatin Tyr	2.37e+02
36	49	52.1	548	3	R14121 Luciferase mutant Tyr	2.37e+02
37	49	52.1	548	1	P94367 Recombinant luciferas	2.37e+02
38	49	52.1	548	16	R84178 Luciferase of Hotaria	2.37e+02
39	49	52.1	548	6	R33709 Thermostable lucifera	2.37e+02
40	49	52.1	548	5	R05788 Luciferase.	2.37e+02
41	49	52.1	550	9	R50011 Firefly luciferase re	2.37e+02
42	49	52.1	743	13	R72801 Vaccinia virus vector	2.37e+02
43	49	52.1	815	5	R28127 Lux::npt-II fusion pr	2.37e+02
44	49	52.1	1498	14	R77085 Hamster sulphonylurea	2.37e+02
45	49	52.1	1582	14	R77087 Rat sulphonylurea rec	2.37e+02

ALIGNMENTS

RESULT 1  
ID R99923 standard; peptide; 12 AA.  
AC R99923;  
DT 22-APR-1997 (first entry)  
DE Osteoclastogenesis inhibitory factor peptide fragment #3.  
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
KW osteoporosis.  
OS Homo sapiens.  
FT Key Location/Qualifiers  
FT Misc\_difference 1  
FT /note= "Any amino acid"  
FN W09626217.A1.  
PD 29-AUG-1996. J00374.  
PF 20-FEB-1996; JP-054977.  
PR 20-FEB-1995; JP-207508.  
PR 21-JUL-1995; JP-207508.  
PA (SNOW) SNOW BRAND MILK PROD CO LTD.  
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
PI WPI: 96-402320/40.  
PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
PT for bone resorption control, esp. treatment of osteoporosis  
PS Claim 1; Page 61; 183pp; Japanese.  
CC The sequences given in R99921-23 and R99926 represent fragments of  
CC the osteoclastogenesis inhibitory factor (OCIF) of the invention.  
CC The OCIF has a molecular weight by SDS-PAGE of 60 kD under reducing  
CC conditions and 120 kD under non-reducing conditions. The protein is  
CC adsorbed onto cation-exchangers or heparin and its activity is  
CC lowered after 10 mins at 70 deg.C or 30 mins at 56 deg.C, and is  
CC lost after 10 mins at 90 deg.C. OCIF is useful in the control of  
CC bone resorption and therefore in the treatment and prevention of  
CC disorders of bone resorption, e.g. osteoporosis.  
SQ Sequence 12 AA;

Query Match 100.0%; Score 94; DB 20; Length 12;  
Best Local Similarity 100.0%; Pred. No. 5.77e-03;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2 irflhsftmyk 12

QY 2 IRLHSFTMYK 12

RESULT 2

R99940 standard; Protein; 326 AA.  
 AC R99940;  
 DT 23-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-DDD1  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 KW osteoporosis.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /note= "Signal peptide"  
 FT Protein 22..326  
 FT /note= "Mature OCIF-DDD2"  
 FT Misc\_difference 198..199  
 FT /note= "Position of deletion, delta 178-252"  
 PN WO9626217-A1.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; J00374.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PR (SNOW) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI; 96-402320/40.  
 DR N-PSDB; T33170.  
 PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 36; Page 113-114; 183pp; Japanese.  
 CC This sequence represents a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC sequence represents OCIF-DDD1 in which amino acids 178-252 of the  
 CC mature OCIF protein are deleted. The OCIF of the invention  
 CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
 CC and 120 kD under non-reducing conditions. The protein is adsorbed onto  
 CC cation-exchangers or heparin and its activity is lowered after 10 mins  
 CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
 CC deg.C. OCIF is useful in the control of bone resorption and therefore  
 CC in the treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 SQ Sequence 326 AA;

Query Match 100.0%; Score 94; DB 20; Length 326;  
 Best Local Similarity 100.0%; Pred. No. 5.77e-03;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 294 irflhsftmyk 304  
 |||||||||  
 2 IRLHSFTMYK 12

RESULT 3  
 ID R99940 standard; Protein; 327 AA.  
 AC R99940;  
 DT 23-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-DDD2  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 KW osteoporosis.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /note= "Signal peptide"  
 FT Protein 22..327  
 FT /note= "Mature OCIF-DDD2"  
 FT Misc\_difference 273..274  
 FT /note= "Position of deletion, delta 253-326"  
 PN WO9626217-A1.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; J00374.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PR (SNOW) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI; 96-402320/40.

Query Match 100.0%; Score 94; DB 20; Length 326;  
 Best Local Similarity 100.0%; Pred. No. 5.77e-03;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 294 irflhsftmyk 304  
 |||||||||  
 2 IRLHSFTMYK 12

DR N-PSDB; T33171.  
 PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 59; Page 115-116; 183pp; Japanese.  
 CC This sequence represents a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC sequence represents OCIF-DDD2 in which amino acids 253-326 of the  
 CC mature OCIF protein are deleted. The OCIF of the invention  
 CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
 CC and 120 kD under non-reducing conditions. The protein is adsorbed onto  
 CC cation-exchangers or heparin and its activity is lowered after 10 mins  
 CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
 CC deg.C. OCIF is useful in the control of bone resorption and therefore  
 CC in the treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 SQ Sequence 327 AA;

Query Match 100.0%; Score 94; DB 20; Length 327;  
 Best Local Similarity 100.0%; Pred. No. 5.77e-03;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 295 irflhsftmyk 305  
 |||||||||  
 2 IRLHSFTMYK 12

RESULT 4  
 ID R99939 standard; Protein; 359 AA.  
 AC R99939;  
 DT 23-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-DCR4  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 KW osteoporosis.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /note= "Signal peptide"  
 FT Protein 22..359  
 FT /note= "Mature OCIF-DCR4"  
 FT Misc\_difference 143..144  
 FT /note= "Position of deletion, delta 123-164"  
 PN WO9626217-A1.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; J00374.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PR (SNOW) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI; 96-402320/40.  
 DR N-PSDB; T33169.

PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 53; Page 111-113; 183pp; Japanese.  
 CC This sequence represents a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC sequence represents OCIF-DCR4 in which amino acids 123-164 of the  
 CC mature OCIF protein are deleted. The OCIF of the invention  
 CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
 CC and 120 kD under non-reducing conditions. The protein is adsorbed onto  
 CC cation-exchangers or heparin and its activity is lowered after 10 mins  
 CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
 CC deg.C. OCIF is useful in the control of bone resorption and therefore  
 CC in the treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 SQ Sequence 359 AA;

Query Match 100.0%; Score 94; DB 20; Length 359;  
 Best Local Similarity 100.0%; Pred. No. 5.77e-03;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 327 irflhsftmyk 337  
 |||||||||

QY 2 IRLHSFTMYK 12

# RESULT 5

ID R99937 standard; Protein: 359 AA.  
 AC R99937; 1997 (first entry)  
 DT 23-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-DCR2.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 OS Osteoporosis.  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /note= "Signal peptide"  
 FT Protein. 22..359  
 FT /note= "Mature OCIF-DCR2"  
 FT Misc.difference 63..64  
 FT /note= "Position of deletion, delta 43-84"  
 PN W09626217-A1.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; J00374.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW ) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI; 96-402320/40.  
 DR N-PSDB; T33167.  
 DT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 47; Page 107-109; 183pp; Japanese.  
 CC This sequence represents a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC sequence represents OCIF-DCR2 in which amino acids 43-84 of the  
 CC mature OCIF protein are deleted. The OCIF of the invention  
 CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
 CC and 120 kD under non-reducing conditions. The protein is adsorbed onto  
 CC cation-exchangers or heparin and its activity is lowered after 10 mins  
 CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
 CC deg.C. OCIF is useful in the control of bone resorption and therefore  
 CC in the treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 SQ Sequence 359 AA;

Query Match 100.0%; Score 94; DB 20; Length 359;  
 Best Local Similarity 100.0%; Pred.No. 5.77e-03;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 327 irflhsftmyk 337

|||||||  
 2 IRLHSFTMYK 12

# RESULT 6

ID R99936 standard; Protein: 360 AA.  
 AC R99936;  
 DT 23-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-DCR1.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 OS Osteoporosis.  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /note= "Signal peptide"  
 FT Protein. 22..360  
 FT /note= "Mature OCIF-DCR1"  
 FT Misc.difference 22..23  
 FT /note= "Position of deletion, delta 2-42"  
 PN W09626217-A1.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; J00374.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.

PA (SNOW ) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI; 96-402320/40.  
 DR N-PSDB; T33166.  
 DT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 44; Page 105-107; 183pp; Japanese.  
 CC This sequence represents a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC sequence represents OCIF-DCR1 in which amino acids 2-42 of the  
 CC mature OCIF protein are deleted. The OCIF of the invention  
 CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
 CC and 120 kD under non-reducing conditions. The protein is adsorbed onto  
 CC cation-exchangers or heparin and its activity is lowered after 10 mins  
 CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
 CC deg.C. OCIF is useful in the control of bone resorption and therefore  
 CC in the treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 SQ Sequence 360 AA;

Query Match 100.0%; Score 94; DB 20; Length 360;  
 Best Local Similarity 100.0%; Pred.No. 5.77e-03;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 328 irflhsftmyk 338

|||||||  
 2 IRLHSFTMYK 12

# RESULT 7

ID R99938 standard; Protein: 360 AA.  
 AC R99938;  
 DT 23-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-DCR3.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 OS Osteoporosis.  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /note= "Signal peptide"  
 FT Protein. 22..360  
 FT /note= "Mature OCIF-DCR3"  
 FT Misc.difference 105..106  
 FT /note= "Position of deletion, delta 85-122"  
 PN W09626217-A1.  
 PD 29-AUG-1996.  
 PF 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW ) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI; 96-402320/40.  
 DR N-PSDB; T33168.  
 DT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 50; Page 109-111; 183pp; Japanese.  
 CC This sequence represents a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC sequence represents OCIF-DCR3 in which amino acids 85-122 of the  
 CC mature OCIF protein are deleted. The OCIF of the invention  
 CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
 CC and 120 kD under non-reducing conditions. The protein is adsorbed onto  
 CC cation-exchangers or heparin and its activity is lowered after 10 mins  
 CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
 CC deg.C. OCIF is useful in the control of bone resorption and therefore  
 CC in the treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 SQ Sequence 360 AA;

Query Match 100.0%; Score 94; DB 20; Length 360;  
 Best Local Similarity 100.0%; Pred.No. 5.77e-03;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 331 irflhsftmyk 341  
 |||||  
 QY 2 IRFLHSFTMYK 12

RESULT 8

ID R99924 standard; Protein; 380 AA.  
 AC R99924;  
 DT 22-APR-1997 (first entry)  
 DE Mature osteoclastogenesis inhibitory factor.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 OS Homo sapiens.  
 PN WO9626217-A1.  
 PD 29-AUG-1996.  
 20-FEB-1996; J00374.  
 20-FEB-1996; JP-054977.  
 21-JUL-1995; JP-207508.  
 (SNOW) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR N-PSDB; T36685.  
 PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 6; Page 62-64; 183pp; Japanese.  
 CC This sequence represents the mature osteoclastogenesis inhibitory  
 factor (OCIF) of the invention. The OCIF has a molecular weight by  
 SDS-PAGE of 60 kD under reducing conditions and 120 kD under non-  
 reducing conditions. The protein is adsorbed onto cation-exchangers  
 or heparin and its activity is lowered after 10 mins at 70 deg.C or  
 30 mins at 56 deg.C, and is lost after 10 mins at 90 deg.C. OCIF is  
 useful in the control of bone resorption and therefore in the  
 treatment and prevention of disorders of bone resorption, e.g.  
 osteoporosis.  
 SQ Sequence 380 AA;

Query Match 100.0%; Score 94; DB 20; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 5.77e-03;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 348 irflhsftmyk 358  
 |||||  
 QY 2 IRFLHSFTMYK 12

RESULT 9

ID R99948 standard; Protein; 393 AA.  
 AC R99948;  
 DT 23-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-CBst.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /note= "Signal peptide"  
 FT Protein 22..399  
 FT /note= "Mature OCIF-CL"  
 PN WO9626217-A1.  
 PD 29-AUG-1996.  
 20-FEB-1996; J00374.  
 20-FEB-1995; JP-054977.  
 21-JUL-1995; JP-207508.  
 (SNOW) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR N-PSDB; T33178.

PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 80; Page 126-128; 183pp; Japanese.  
 CC This sequence represents a mutated version of the full length  
 osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 sequence represents OCIF-CBst in which Gln 371 is substituted by  
 Leu and amino acids 373-380 of the mature OCIF protein are deleted.  
 CC These changes are caused by the introduction of a restriction site in  
 the DNA encoding this protein. The OCIF of the invention has a  
 molecular weight by SDS-PAGE of 60 kD under reducing conditions  
 and 120 kD under non-reducing conditions. The protein is adsorbed onto  
 cation-exchangers or heparin and its activity is lowered after 10 mins  
 at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
 deg.C. OCIF is useful in the control of bone resorption and therefore  
 in the treatment and prevention of disorders of bone resorption, e.g.  
 osteoporosis.  
 SQ Sequence 393 AA;

Query Match 100.0%; Score 94; DB 20; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 5.77e-03;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 369 irflhsftmyk 379  
 |||||  
 QY 2 IRFLHSFTMYK 12

RESULT 10

ID R99942 standard; Protein; 399 AA.  
 AC R99942;  
 DT 23-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-CL.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /note= "Signal peptide"  
 FT Protein 22..399  
 FT /note= "Mature OCIF-CL"  
 PN WO9626217-A1.  
 PD 29-AUG-1996.  
 20-FEB-1996; J00374.  
 20-FEB-1995; JP-054977.  
 21-JUL-1995; JP-207508.  
 (SNOW) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR N-PSDB; T33172.

PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 62; Page 117-119; 183pp; Japanese.  
 CC This sequence represents a mutated version of the full length  
 osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 sequence represents OCIF-CL in which amino acids 379-380 of the  
 mature OCIF protein are deleted. The OCIF of the invention  
 has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
 and 120 kD under non-reducing conditions. The protein is adsorbed onto  
 cation-exchangers or heparin and its activity is lowered after 10 mins  
 at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
 deg.C. OCIF is useful in the control of bone resorption and therefore  
 in the treatment and prevention of disorders of bone resorption, e.g.  
 osteoporosis.  
 SQ Sequence 399 AA;

Query Match 100.0%; Score 94; DB 20; Length 399;  
 Best Local Similarity 100.0%; Pred. No. 5.77e-03;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 369 irflhsftmyk 379  
 |||||  
 QY 2 IRFLHSFTMYK 12



```

RESULT 11
ID R99934 standard; Protein; 401 AA.
AC R99934;
DT 22-APR-1997 (first entry)
DE Mutated OCIF, OCIF-C22S.
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;
OS Osteoporosis.
FH Key Location/Qualifiers
FT Peptide 1..21
FT /note= "Signal peptide"
FT Protein 22..401
FT /note= "Mature OCIF-C22S"
FT Misc_difference 277
FT /label= C22S
PN WO9626217-A1.
PD 29-AUG-1996.
PF 20-FEB-1996; J00374.
PR 20-FEB-1995; JP-054977.
PS 21-JUL-1995; JP-207508.
(SNOW ) SNOW BRAND MILK PROD CO LTD.
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;
DR N-PSDB; T33162.
DR DNA encoding osteoclastogenesis inhibitory factor protein - useful
PT for bone resorption control, esp. treatment of osteoporosis
PS Claim 38; Page 100-102; 183pp; Japanese.
CC This sequence represents a mutated version of the full length
CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This
CC sequence represents OCIF-C22S in which the 22nd Cys residue in the
CC mature OCIF protein is substituted by Ser. The OCIF of the invention
CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions
CC and 120 kD under non-reducing conditions. The protein is adsorbed onto
CC cation-exchangers or heparin and its activity is lowered after 10 mins
CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90
CC deg.C. OCIF is useful in the control of bone resorption and therefore
CC in the treatment and prevention of disorders of bone resorption, e.g.
CC osteoporosis.
SQ Sequence 401 AA;

Query Match 100.0%; Score 94; DB 20; Length 401;
Best Local Similarity 100.0%; Pred. No. 5.77e-03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 369 lrflhsftmyk 379
|||||
2 lrflhsftmyk 12

RESULT 12
ID R99932 standard; Protein; 401 AA.
AC R99932;
DT 22-APR-1997 (first entry)
DE Mutated OCIF, OCIF-C20S.
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;
OS Osteoporosis.
FH Key Location/Qualifiers
FT Peptide 1..21
FT /note= "Signal peptide"
FT Protein 22..401
FT /note= "Mature OCIF-C20S"
FT Misc_difference 202
FT /label= C20S
PN WO9626217-A1.
PD 29-AUG-1996.
PF 20-FEB-1996; J00374.
PR 20-FEB-1995; JP-054977.
PS 21-JUL-1995; JP-207508.
(SNOW ) SNOW BRAND MILK PROD CO LTD.
PA Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;
PA Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;
DR N-PSDB; T33165.
DR DNA encoding osteoclastogenesis inhibitory factor protein - useful
PT for bone resorption control, esp. treatment of osteoporosis
PS Claim 41; Page 103-105; 183pp; Japanese.
CC This sequence represents a mutated version of the full length
CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This
CC sequence represents OCIF-C23S in which the 23rd Cys residue in the
CC mature OCIF protein is substituted by Ser. The OCIF of the invention
CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions
CC and 120 kD under non-reducing conditions. The protein is adsorbed onto
CC cation-exchangers or heparin and its activity is lowered after 10 mins
CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90
CC deg.C. OCIF is useful in the control of bone resorption and therefore
CC in the treatment and prevention of disorders of bone resorption, e.g.
CC osteoporosis.
SQ Sequence 401 AA;

Query Match 100.0%; Score 94; DB 20; Length 401;
Best Local Similarity 100.0%; Pred. No. 5.77e-03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 369 lrflhsftmyk 379
|||||
2 lrflhsftmyk 12

```

```

PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;
DR N-PSDB; T33162.
DR DNA encoding osteoclastogenesis inhibitory factor protein - useful
PT for bone resorption control, esp. treatment of osteoporosis
PS Claim 32; Page 96-98; 183pp; Japanese.
CC This sequence represents a mutated version of the full length
CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This
CC sequence represents OCIF-C20S in which the 20th Cys residue in the
CC mature OCIF protein is substituted by Ser. The OCIF of the invention
CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions
CC and 120 kD under non-reducing conditions. The protein is adsorbed onto
CC cation-exchangers or heparin and its activity is lowered after 10 mins
CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90
CC deg.C. OCIF is useful in the control of bone resorption and therefore
CC in the treatment and prevention of disorders of bone resorption, e.g.
CC osteoporosis.
SQ Sequence 401 AA;

Query Match 100.0%; Score 94; DB 20; Length 401;
Best Local Similarity 100.0%; Pred. No. 5.77e-03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 369 lrflhsftmyk 379
|||||
2 lrflhsftmyk 12

RESULT 13
ID R99935 standard; Protein; 401 AA.
AC R99935;
DT 22-APR-1997 (first entry)
DE Mutated OCIF, OCIF-C23S.
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;
OS Osteoporosis.
FH Key Location/Qualifiers
FT Peptide 1..21
FT /note= "Signal peptide"
FT Protein 22..401
FT /note= "Mature OCIF-C23S"
FT Misc_difference 400
FT /label= C23S
PN WO9626217-A1.
PD 29-AUG-1996.
PF 20-FEB-1996; J00374.
PR 20-FEB-1995; JP-054977.
PS 21-JUL-1995; JP-207508.
(SNOW ) SNOW BRAND MILK PROD CO LTD.
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;
DR N-PSDB; T33165.
DR DNA encoding osteoclastogenesis inhibitory factor protein - useful
PT for bone resorption control, esp. treatment of osteoporosis
PS Claim 41; Page 103-105; 183pp; Japanese.
CC This sequence represents a mutated version of the full length
CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This
CC sequence represents OCIF-C23S in which the 23rd Cys residue in the
CC mature OCIF protein is substituted by Ser. The OCIF of the invention
CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions
CC and 120 kD under non-reducing conditions. The protein is adsorbed onto
CC cation-exchangers or heparin and its activity is lowered after 10 mins
CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90
CC deg.C. OCIF is useful in the control of bone resorption and therefore
CC in the treatment and prevention of disorders of bone resorption, e.g.
CC osteoporosis.
SQ Sequence 401 AA;

Query Match 100.0%; Score 94; DB 20; Length 401;
Best Local Similarity 100.0%; Pred. No. 5.77e-03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 369 lrflhsftmyk 379
|||||
2 lrflhsftmyk 12

```

```

Query Match 100.0%; Score 94; DB 20; Length 401;
Best Local Similarity 100.0%; Pred. No. 5.77e-03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Db 369 irflhsftmyk 379  
|||||  
QY 2 IRLHSTMYK 12

## RESULT 14

ID R99931 standard; Protein; 401 AA.  
AC R99931;  
DT 22-APR-1997 (first entry)  
DE Mutated OCIF, OCIF-C19S.  
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
OS Synthetic.  
FH Key Location/Qualifiers  
FT Peptide 1..21  
FT /note= "Signal peptide"  
FT Protein 22..401  
FT /label= C19S  
FT /note= "Mature OCIF-C19S"  
FT /misc\_difference 195  
PN WO9626217-A1.  
PD 29-AUG-1996.  
PF 20-FEB-1996; J00374.  
PR 20-FEB-1995; JP-054977.  
PR 21-JUL-1995; JP-207508.  
PA (SNOW ) SNOW BRAND MILK PROD CO LTD.  
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
DR WPI: 96-402320/40.  
DR N-PSDB: T33161.  
PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
for bone resorption control, esp. treatment of osteoporosis  
Claim 29; Page 94-96; 183pp; Japanese.  
CC This sequence represents a mutated version of the full length  
osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
sequence represents OCIF-C19S in which the 19th Cys residue in the  
mature OCIF protein is substituted by Ser. The OCIF of the invention  
has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
and 120 kD under non-reducing conditions. The protein is adsorbed onto  
cation-exchangers or heparin and its activity is lowered after 10 mins  
at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
deg.C. OCIF is useful in the control of bone resorption and therefore  
in the treatment and prevention of disorders of bone resorption, e.g.  
osteoporosis.  
SQ Sequence 401 AA;

Query Match 100.0%; Score 94; DB 20; Length 401;  
Best Local Similarity 100.0%; Pred. No. 5.77e-03;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 369 irflhsftmyk 379  
|||||  
QY 2 IRLHSTMYK 12

## RESULT 15

ID R99925 standard; Protein; 401 AA.  
AC R99925;  
DT 22-APR-1997 (first entry)  
DE Full length osteoclastogenesis inhibitory factor.  
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT Peptide 1..21  
FT /note= "Signal peptide"  
FT Protein 22..401  
FT /note= "Mature OCIF, claim 6"  
PN WO9626217-A1.  
PD 29-AUG-1996.  
PF 20-FEB-1996; J00374.  
PR 20-FEB-1995; JP-054977.

PR 21-JUL-1995; JP-207508.  
PA (SNOW ) SNOW BRAND MILK PROD CO LTD.  
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
DR WPI: 96-402320/40.  
DR N-PSDB: T36685.  
PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
for bone resorption control, esp. treatment of osteoporosis  
PS Disclosure; Page 64-66; 183pp; Japanese.  
CC This sequence represents the full length osteoclastogenesis inhibitory  
factor (OCIF) of the invention. The OCIF has a molecular weight by  
SDS-PAGE of 60 kD under reducing conditions and 120 kD under non-  
reducing conditions. The protein is adsorbed onto cation-exchangers  
or heparin and its activity is lowered after 10 mins at 70 deg.C or  
30 mins at 56 deg.C, and is lost after 10 mins at 90 deg.C. OCIF is  
useful in the control of bone resorption and therefore in the  
treatment and prevention of disorders of bone resorption, e.g.  
osteoporosis.  
SQ Sequence 401 AA;

Query Match 100.0%; Score 94; DB 20; Length 401;  
Best Local Similarity 100.0%; Pred. No. 5.77e-03;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 369 irflhsftmyk 379  
|||||  
QY 2 IRLHSTMYK 12

Search completed: Tue Dec 2 14:46:45 1997  
Job time : 11 secs.

\*\*\*\*\*  
MORF  
\*\*\*\*\*  
(TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Dec 2 14:46:05 1997: MasPar time 2.71 Seconds  
128.122 Million cell updates/sec

ular output not generated.

File: >US-08-915-004-3  
Description: (1-12) from US08915004.pep  
Perfect Score: 94  
Sequence: 1 XIRFLHSFTMYK 12

Scoring table: PAM 150  
Gap 15

Searched: 91006 seqs, 28888923 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pir51  
1:ann1 2:ann2 3:ann3 4:ann4 5:unann1 6:unann2 7:unann3  
8:unann4 9:unann5 10:unann6 11:unann7 12:unann8  
13:unann9 14:unann10 15:unenc 16:unrev

Statistics: Mean 25.879; Variance 44.392; scale 0.583

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description	Pred. No.
1	59	62.8	1441	4	GNVULC	6.93e+00
2	58	61.7	67	8	A-ORF-S protein - va	9.85e+00
3	56	59.6	446	8	beta-glucosidase [EC	1.97e+01
4	56	59.6	1160	7	paraspore crystal p	1.97e+01
5	56	59.6	1165	7	paraspore crystal p	1.97e+01
6	55	58.5	322	10	hypothetical protein	2.77e+01
7	55	58.5	785	16	hypothetical protein	2.77e+01
8	54	57.4	325	13	melanocortin recepto	3.89e+01
9	54	57.4	772	8	SR protein - human h	3.89e+01
10	54	57.4	849	4	noncapsid protein NS	3.89e+01
11	54	57.4	941	8	hypothetical protein	3.89e+01
12	53	56.4	231	8	C20L protein - vario	5.42e+01
13	53	56.4	231	8	F16L protein - vacci	5.42e+01
14	53	56.4	838	9	nitrate reductase (N	5.42e+01
15	53	56.4	839	9	nitrite reductase ni	5.42e+01
16	53	56.4	1522	12	sodium channel prote	5.42e+01
17	53	56.4	2711	16	faf-fat facets gene	5.42e+01
18	53	56.4	2747	16	faf-fat facets gene	5.42e+01
19	52	55.3	155	16	recx protein - pseu	7.52e+01
20	52	55.3	232	9	rfay protein - Esche	7.52e+01
21	52	55.3	232	9	Rfay - Escherichia c	7.52e+01

22 52 55.3 281 16 S67440 hypothetical protein 7.52e+01  
23 52 55.3 327 16 S11944 QUTG protein - Emeri 7.52e+01  
24 52 55.3 330 12 S08500 QUTG protein - Emeri 7.52e+01  
25 52 55.3 468 12 S44815 F44B9.4 protein - Ca 7.52e+01  
26 52 55.3 515 5 A32931 glucosylceramidase ( 7.52e+01  
27 52 55.3 578 6 A34283 H+-transporting ATP 7.52e+01  
28 52 55.3 676 8 F36843 ATP/GTP-binding prot 7.52e+01  
29 52 55.3 676 8 D42511 I8R protein - vacci 7.52e+01  
30 52 55.3 676 4 WZV418 I8 protein - vaccini 7.52e+01  
31 52 55.3 702 12 A34434 arylphorin alpha cha 7.52e+01  
32 52 55.3 1351 11 S56218 DNA-directed RNA pol 7.52e+01  
33 52 55.3 2161 1 A45389 RNA-directed RNA pol 7.52e+01  
34 51 54.3 350 10 S34557 hypothetical protein 1.04e+02  
35 51 54.3 427 14 I49603 transcription regula 1.04e+02  
36 51 54.3 537 4 F0MVW7 gag polyprotein - ba 1.04e+02  
37 51 54.3 586 4 TVVEPP large T antigen - bo 1.04e+02  
38 51 54.3 642 8 I40071 threonyl-tRNA synthe 1.04e+02  
39 51 54.3 659 14 A35723 70K peroxisomal memb 1.04e+02  
40 51 54.3 659 13 S20313 peroxisomal membrane 1.04e+02  
41 51 54.3 703 12 B34434 arylphorin beta chai 1.04e+02  
42 51 54.3 718 9 B64040 hypothetical protein 1.04e+02  
43 51 54.3 760 5 A40195 meprin A (EC 3.4.24. 1.04e+02  
44 51 54.3 882 14 A42855 N-heparan sulfate su 1.04e+02  
45 51 54.3 1023 12 A47236 thiazide-sensitive e 1.04e+02

ALIGNMENTS

RESULT 1  
ENTRY GNVULC #type complete  
TITLE M polyprotein precursor - La Crosse virus (isolate 74-32813)  
CONTAINS glycoprotein G1; glycoprotein G2; nonstructural protein NS-M  
ORGANISM #formal\_name La Crosse virus  
DATE 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 05-Jan-1996  
ACCESSIONS A29377  
REFERENCE A29377  
#authors Grady, L.J.; Sanders, M.L.; Campbell, W.P.  
#journal J. Gen. Virol. (1987) 68:3057-3071  
#title The sequence of the M RNA of an isolate of La Crosse virus.  
#cross-references MUID:88089508  
#accession A29377  
#molecule\_type mRNA  
#residues 1-1441 #label GRA  
#note the authors translated the codon GCC for residue 164 as Gly

COMMENT Specific enzymatic cleavages in vivo yield mature proteins including nonstructural protein NS-M, glycoprotein G1, and glycoprotein G2. However, exact cleavage sites are undetermined.

GENETICS  
#map\_position segment M  
CLASSIFICATION #superfamily bunyavirus: M polyprotein  
KEYWORDS glycoprotein; nonstructural protein; polyprotein  
FEATURE  
1-13 #domain signal sequence #status predicted #label SIG\  
14-1441 #product M polyprotein #status predicted #label MPP\  
30,57,245,490, #binding\_site carbohydrate (Asn) (covalent) #status predicted  
1177 #length 1441 #molecular\_weight 162540 #checksum 6648  
SUMMARY

Query Match 62.8%; Score 59; DB 4; Length 1441;  
Best Local Similarity 55.6%; Pred. No. 6.93e+00;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 388 rflniyamy 396  
QY 3 RFLHSFTMY 11

RESULT 2  
ENTRY #type complete  
TITLE A-ORF-S protein - vaccinia virus (strains WR and Copenhagen)

```

QY      2  IRLHSTMY 11

RESULT      4
ENTRY      S32647      #type complete
TITLE      parasporal crystal protein - Bacillus thuringiensis
ORGANISM   #formal_name Bacillus thuringiensis
DATE       20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
12-Apr-1995
ACCESSIONS S32647
REFERENCE   S32645
#authors   Lambert, B.
#submission submitted to the EMBL Data Library, April 1993
#accession S32647
##status   preliminary
##molecule_type DNA
##residues 1-l1160 ##label LAM
##cross-references EMBL:222511
CLASSIFICATION #superfamily parasporal crystal protein
SUMMARY     #length 1160 #molecular-weight 130968 #checksum 7196

Query Match      59.6% Score 56; DB 7; Length 1160;
Best Local Similarity 60.0% Pred. No. 1.97e+01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 300 vdfinsftiy 309
: ||:||||:|
QY 2  IRLHSTMY 11

RESULT      5
ENTRY      S11446      #type complete
TITLE      parasporal crystal protein cry7D - Bacillus thuringiensis
ORGANISM   #formal_name Bacillus thuringiensis
DATE       21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change
03-May-1996
ACCESSIONS S11446
REFERENCE   S11446
#authors   Hoefte, H.; Soetaert, P.; Jansens, S.; Peferoen, M.
#journal   Nucleic Acids Res. (1990) 18:3545
#title     Nucleotide sequence and deduced amino acid sequence of a new
Leptodiptera-specific crystal protein gene from Bacillus
thuringiensis.
#cross-references M01D:91016842
#accession S11446
##status   preliminary
##molecule_type DNA
##residues 1-l1165 ##label HOE
##cross-references EMBL:X54160
CLASSIFICATION #superfamily parasporal crystal protein
SUMMARY     #length 1165 #molecular-weight 132480 #checksum 8294

Query Match      59.0% Score 56; DB 7; Length 1165;
Best Local Similarity 60.0% Pred. No. 1.97e+01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 300 vdfinsftiy 309
: ||:||||:|
QY 2  IRLHSTMY 11

RESULT      6
ENTRY      G64388      #type complete
TITLE      hypothetical protein MJ0711 - Methanococcus jannaschii
ORGANISM   #formal_name Methanococcus jannaschii
DATE       13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change
13-Sep-1996
ACCESSIONS G64388
REFERENCE   A64300
#authors   Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann,
R.D.; Sutton, G.G.; Blake, J.A.; Fitzgerald, L.M.; Clayton,
R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.;
Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.;

```

Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;  
 Scott, J.L.; Geoghagen, N.S.M.; Weidman, J.F.; Fuhrmann,  
 J.L.; Nguyen, D.; Utterback, T.R.; Kelley, J.M.; Peterson,  
 J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts,  
 K.M.; Hurst, M.A.; Kane, B.P.; Borodovsky, M.; Klensk,  
 H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.  
 Science (1996) 273:1058-1073  
 Complete genome sequence of the methanogenic archaeon,  
 Methanococcus jannaschii.

accession G64368  
 #status preliminary; nucleic acid sequence not shown;  
 translation not shown

##molecule\_type DNA  
 ##residues 1-322 #label BUL  
 ##cross-references GB:L77117; TIGR:MJ0711; CDS\_PID:g1510799

GENETICS

##map\_position FOR643997-644965

SUMMARY

length 322 #molecular-weight 38631 #checksum 3071

Query Match 58.5%; Score 55; DB 10; Length 322;

Best Local Similarity 55.6%; Pred. No. 2.77e+01;

Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 104 kflesytl112

QY 3 RFLHSFTMY 11

RESULT 7

ENTRY S63652 #type complete

TITLE hypothetical protein 785 - Allomyces macrogynus mitochondrion

ORGANISM #formal\_name mitochondrion Allomyces macrogynus

DATE 28-Oct-1996 #sequence\_revision 28-Oct-1996 #text\_change

28-Oct-1996

accession S63652

REFERENCE S63635

#authors Paquin, B.; Lang, B.F.

#journal J. Mol. Biol. (1996) 255:688-701

#title The mitochondrial DNA of Allomyces macrogynus: the complete

genomic sequence from an ancestral fungus.

accession S63652

##status preliminary

##residues 1-785 #label PAQ

##cross-references EMBL:U41288

SUMMARY

length 785 #molecular-weight 88119 #checksum 2052

Query Match 58.5%; Score 55; DB 16; Length 785;

Best Local Similarity 44.4%; Pred. No. 2.77e+01;

Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 400 ylhafdmf 408

QY 4 FLHSFTMY 12

RESULT 8

ENTRY JN0764 #type complete

TITLE melanocortin receptor 2 - human

ORGANISM #formal\_name Homo sapiens #common\_name man

DATE 14-Jul-1994 #sequence\_revision 14-Jul-1994 #text\_change

23-Aug-1996

accession JN0764; S37042; S40681

REFERENCE S40681

#authors Chhajlani, V.; Muceniece, R.; Wikberg, J.E.S.

#journal Biochem. Biophys. Res. Commun. (1993) 195:866-873

#title Molecular cloning of a novel human melanocortin receptor.

accession JN0764

##molecule\_type DNA

##residues 1-325 #label CHH

##cross-references EMBL:225470

##experimental\_source brain

KEYWORDS

glycoprotein; receptor; transmembrane protein

FEATURE

38-61  
 74-97  
 115-138  
 156-179  
 187-211  
 240-265  
 274-297  
 2,20,28

#domain transmembrane #status predicted #label TM1\

#domain transmembrane #status predicted #label TM2\

#domain transmembrane #status predicted #label TM3\

#domain transmembrane #status predicted #label TM4\

#domain transmembrane #status predicted #label TM5\

#domain transmembrane #status predicted #label TM6\

#domain transmembrane #status predicted #label TM7\

#binding site carbohydrate (Asn) (covalent) #status

predicted

SUMMARY

length 325 #molecular-weight 36618 #checksum 589

Query Match 57.4%; Score 54; DB 13; Length 325;

Best Local Similarity 55.6%; Pred. No. 3.89e+01;

Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 272 rfmshfmy 280

QY 3 RFLHSFTMY 11

RESULT 9

ENTRY F33560 #type complete

TITLE SR protein - human herpesvirus 6 (strain Uganda-1102)

ORGANISM #formal\_name human herpesvirus 6

DATE 26-Jul-1991 #sequence\_revision 26-Jul-1991 #text\_change

08-Apr-1994

accession F33560

REFERENCE A33560

#authors Lawrence, G.L.; Chee, M.; Craxton, M.A.; Gompels, U.A.;

Honess, R.W.; Barrell, B.G.

#journal J. Virol. (1990) 64:287-299

#title Human herpesvirus 6 is closely related to human

cytomegalovirus.

#cross-references MUID:90080132

accession F33560

##status preliminary

##molecule\_type DNA

##residues 1-772 #label LAW

##cross-references GB:M28243

SUMMARY

length 772 #molecular-weight 88746 #checksum 7872

Query Match 57.4%; Score 54; DB 8; Length 772;

Best Local Similarity 85.7%; Pred. No. 3.89e+01;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 470 hsfvmyk 476

QY 6 HSFTMYK 12

RESULT 10

ENTRY UYPVAD #type complete

TITLE noncapsid protein NS1 - Aedes densovirus (strain GKV 002 002)

ORGANISM #formal\_name Aedes densovirus

DATE host Aedes aegypti (yellow fever mosquito)

30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change

08-Apr-1994

accession A40784

REFERENCE A40784

#authors Afanasiev, B.N.; Galyov, E.E.; Buchatsky, L.P.; Kozlov, Y.V.

#journal Virology (1991) 185:323-336

#title Nucleotide sequence and genomic organization of Aedes

densovirus.

#cross-references MUID:92024090

accession A40784

##molecule\_type DNA

##residues 1-849 #label AFA

##cross-references GB:M37899

CLASSIFICATION #superfamily Aedes densovirus noncapsid protein NS1

KEYWORDS noncapsid protein

SUMMARY

length 849 #molecular-weight 97542 #checksum 1776

Query Match 57.4%; Score 54; DB 4; Length 849;

```

Best Local Similarity 30.0%; Pred. No. 3.89e+01;
Matches 3; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Db 416 vxflnlyily 425
      ::||::: |
QY 2 IRLHSFTMY 11

RESULT 11
ENTRY
TITLE
ORGANISM
#note
DATE
ACCESSIONS
REFERENCE
#authors
S09851 #type complete
S09749 hypothetical protein UL87 - human cytomegalovirus (strain
AD169)
#formal_name human cytomegalovirus, human herpesvirus 5
#host Homo sapiens (man)
#07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change
27-Jan-1995
S09851
S09749
Chee, M.S.; Bankier, A.T.; Beck, S.; Bohni, R.; Brown, C.M.;
Cerny, R.; Horsnell, T.; Hutchison III, C.A.; Kozzarides,
T.; Martignetti, J.A.; Preddie, E.; Satchwell, S.C.;
Tomlinson, P.; Weston, K.M.; Barrell, B.G.
Curr. Top. Microbiol. Immunol. (1990) 154:125-169
Analysis of the protein-coding content of the sequence of
human cytomegalovirus strain AD169.
#cross-references MUID:90269039
#accession S09851
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-941 #label CHE
#cross-references EMBL:X17403
#note this sequence was submitted to the EMBL Data Library,
December 1989
SUMMARY
#length 941 #molecular-weight 104803 #checksum 1180
Query Match 57.4%; Score 54; DB 8; Length 941;
Best Local Similarity 85.7%; Pred. No. 3.89e+01;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 594 hfvmymk 600
      ||| |||
QY 6 HSFTMYK 12

RESULT 12
ENTRY
TITLE
ORGANISM
#note
DATE
ACCESSIONS
REFERENCE
#authors
B36841 #type complete
B36859 C20L protein - variola virus (strain India-1967)
#formal_name variola virus
#30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
30-Sep-1993
B36841
B36859
Blinov, V.M.
#submitted to GenBank, November 1992
#description not shown.
#accession B36841
#status preliminary
#molecule_type DNA
#residues 1-231 #label BLI
#cross-references GB:X69198
SUMMARY
#length 231 #molecular-weight 26497 #checksum 4488
Query Match 56.4%; Score 53; DB 8; Length 231;
Best Local Similarity 50.0%; Pred. No. 5.42e+01;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 198 rfvnklmkyk 207
      ||::: |||
QY 3 RFLHSFTMYK 12

RESULT 13
ENTRY
TITLE
ORGANISM
#note
DATE
ACCESSIONS
REFERENCE
#authors
C42508 #type complete
#formal_name vaccinia virus (strain Copenhagen)
#host Homo sapiens (man)
#09-Nov-1990 #sequence_revision 09-Nov-1990 #text_change
08-Apr-1994
C42508
A33172
Johnson, G.P.
#submitted to GenBank, June 1990
#accession C42508
#status preliminary
#molecule_type DNA
#residues 1-231 #label JOH
SUMMARY
#length 231 #molecular-weight 26605 #checksum 7005
Query Match 56.4%; Score 53; DB 8; Length 231;
Best Local Similarity 50.0%; Pred. No. 5.42e+01;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 198 rfvnklmkyk 207
      ||::: |||
QY 3 RFLHSFTMYK 12

RESULT 14
ENTRY
TITLE
ORGANISM
#note
DATE
ACCESSIONS
REFERENCE
#authors
S04349 #type complete
S04349 nitrate reductase (NADH) (EC 1.6.6.1) - Escherichia coli
#formal_name Escherichia coli
#15-Oct-1994 #sequence_revision 15-Oct-1994 #text_change
18-Nov-1994
S04349
S04349
Bell, A.I.; Gaston, K.L.; Cole, J.A.; Busby, S.J.W.
Nucleic Acids Res. (1989) 17:3865-3874
Cloning of binding sequences for the Escherichia coli
transcription activators, FNR and CRP: location of bases
involved in discrimination between FNR and CRP.
#cross-references MUID:89282391
#accession S04349
#status translation not shown
#molecule_type DNA
#residues 1-838 #label BEL
#cross-references EMBL:X14202
GENETICS
#gene nirB
KEYWORDS
oxidoreductase
SUMMARY
#length 838 #molecular-weight 92717 #checksum 1768
Query Match 56.4%; Score 53; DB 9; Length 838;
Best Local Similarity 40.0%; Pred. No. 5.42e+01;
Matches 4; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 729 klyldrfmmf 738
      ||::: |
QY 2 IRLHSFTMY 11

RESULT 15
ENTRY
TITLE
ORGANISM
#note
DATE
ACCESSIONS
REFERENCE
#authors
S10791 #type complete
S10791 nitrite reductase nirB - Escherichia coli
#formal_name Escherichia coli
#21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change
10-Nov-1995
S10791
S10791
Peakman, T.; Crouzet, J.; Mayaux, J.F.; Busby, S.; Mohan, S.;
Harborne, N.; Wootton, J.; Nicolson, R.; Cole, J.
Eur. J. Biochem. (1990) 191:315-323
Nucleotide sequence, organisation and structural analysis of
the products of genes in the nirB-cysg region of the
Escherichia coli K-12 chromosome.
#cross-references MUID:90345936
#accession S10791

```

```

TITLE
ORGANISM
#note
DATE
ACCESSIONS
REFERENCE
#authors
#submission
#accession
#status
#molecule_type
#residues
SUMMARY
Query Match 56.4%; Score 53; DB 8; Length 231;
Best Local Similarity 50.0%; Pred. No. 5.42e+01;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 198 rfvnklmkyk 207
      ||::: |||
QY 3 RFLHSFTMYK 12

RESULT 14
ENTRY
TITLE
ORGANISM
#note
DATE
ACCESSIONS
REFERENCE
#authors
S04349 #type complete
S04349 nitrate reductase (NADH) (EC 1.6.6.1) - Escherichia coli
#formal_name Escherichia coli
#15-Oct-1994 #sequence_revision 15-Oct-1994 #text_change
18-Nov-1994
S04349
S04349
Bell, A.I.; Gaston, K.L.; Cole, J.A.; Busby, S.J.W.
Nucleic Acids Res. (1989) 17:3865-3874
Cloning of binding sequences for the Escherichia coli
transcription activators, FNR and CRP: location of bases
involved in discrimination between FNR and CRP.
#cross-references MUID:89282391
#accession S04349
#status translation not shown
#molecule_type DNA
#residues 1-838 #label BEL
#cross-references EMBL:X14202
GENETICS
#gene nirB
KEYWORDS
oxidoreductase
SUMMARY
#length 838 #molecular-weight 92717 #checksum 1768
Query Match 56.4%; Score 53; DB 9; Length 838;
Best Local Similarity 40.0%; Pred. No. 5.42e+01;
Matches 4; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 729 klyldrfmmf 738
      ||::: |
QY 2 IRLHSFTMY 11

RESULT 15
ENTRY
TITLE
ORGANISM
#note
DATE
ACCESSIONS
REFERENCE
#authors
S10791 #type complete
S10791 nitrite reductase nirB - Escherichia coli
#formal_name Escherichia coli
#21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change
10-Nov-1995
S10791
S10791
Peakman, T.; Crouzet, J.; Mayaux, J.F.; Busby, S.; Mohan, S.;
Harborne, N.; Wootton, J.; Nicolson, R.; Cole, J.
Eur. J. Biochem. (1990) 191:315-323
Nucleotide sequence, organisation and structural analysis of
the products of genes in the nirB-cysg region of the
Escherichia coli K-12 chromosome.
#cross-references MUID:90345936
#accession S10791

```

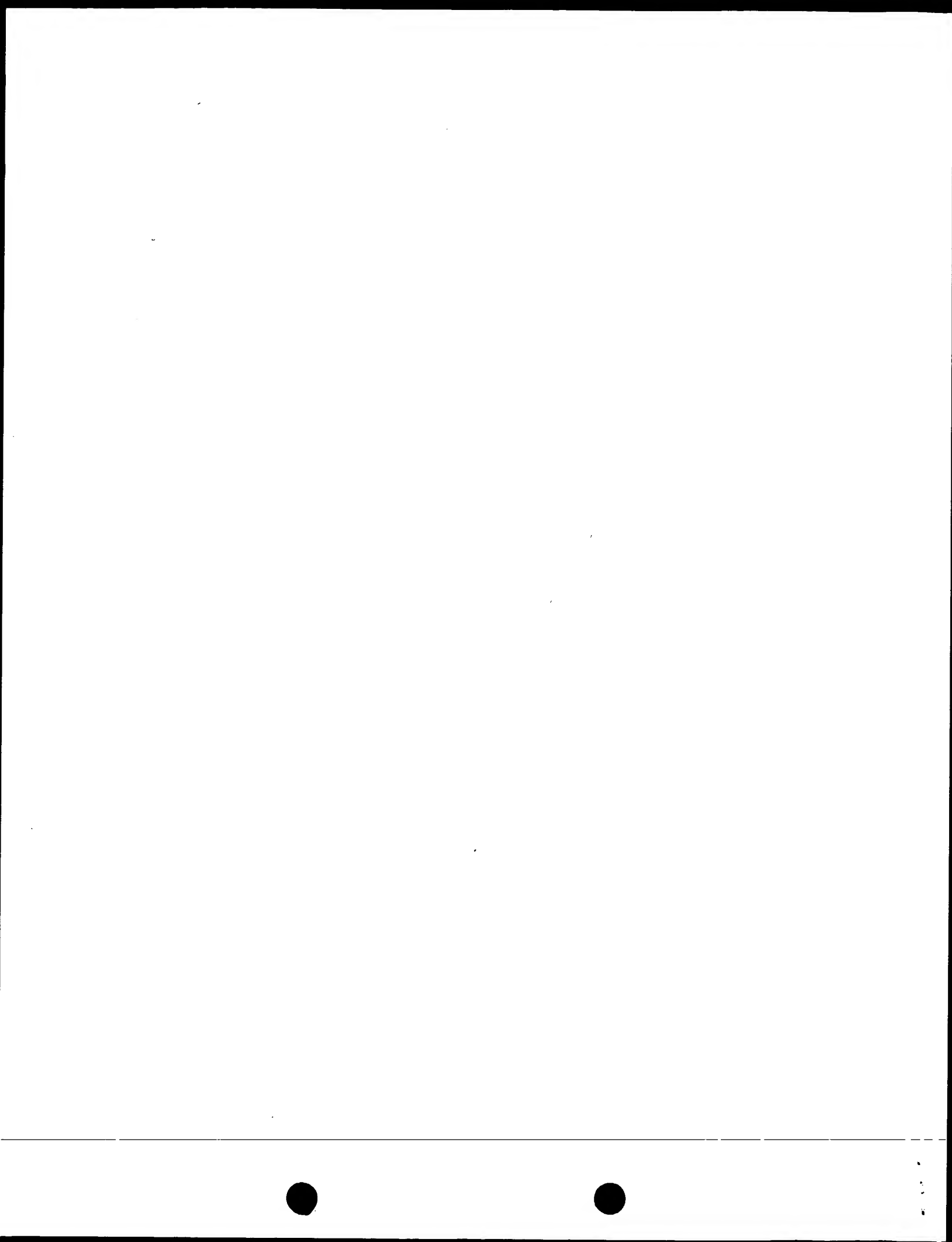
```

##molecule_type DNA
##residues 1-839 ##label PEA
GENETICS
#gene nlrB
SUMMARY #length 839 #molecular-weight 92121 #checksum 2150
Query Match 56.4%; Score 53; DB 9; Length 839;
Best Local Similarity 40.0%; Pred.No. 5.42e+01;
Matches 4; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 730 ikylidrfmmf 739
|::| | |
QY 2 IRLHSEPTMY 11

```

Search completed: Tue Dec 2 14:46:16 1997  
Job time : 11 secs.





\*\*\*\*\*

(TM)

\*\*\*\*\*

Release 2.1D John F. Collins, BioComputing Research Unit.  
 Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
 Distribution rights by Intelligenetics, Inc.

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Dec 2 14:45:38 1997; MasPar time 1.94 Seconds  
 Molecular output not generated. 130.901 Million cell updates/sec

Title: >US-08-915-004-3  
 Description: (1-12) from US08915004.pep  
 Perfect Score: 94  
 Sequence: 1 XIRFLHSFMYK 12

Scoring table: PAM 150  
 Gap 15

Searched: 59021 seqs, 21210388 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: swiss-prot34  
 1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
 8:part8 9:part9 10:part10 11:part11

Statistics: Mean 26.933; Variance 36.903; scale 0.730

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	59	62.8	1441	10	VGLM_BUNL7 M POLYPROTEIN PRECURS	1.41e+00
2	58	61.7	67	11	YVAS_VACCV HYPOTHETICAL 7.5 KD P	2.14e+00
3	58	61.7	790	11	YP84_CAEEL HYPOTHETICAL 90.9 KD	2.14e+00
4	57	60.6	174	11	YRKL_BACSU HYPOTHETICAL NAD(P)H	3.26e+00
5	56	59.6	446	1	BGLA_THEME BETA-GLUCOSIDASE A (E	4.92e+00
6	56	59.6	1165	2	CRYS_BACTA 132 KD CRYSTAL PROTEI	4.92e+00
7	54	57.4	325	6	MC5R_HUMAN MELANOCORTIN-5 RECEPT	1.10e+01
8	54	57.4	772	10	UL87_HSV6U PROTEIN U58.	1.10e+01
9	54	57.4	775	10	UL87_HSV7J PROTEIN U58.	1.10e+01
10	54	57.4	849	10	VNCS_AEDEV NONCAPSID PROTEIN NS-	1.10e+01
11	54	57.4	941	10	UL87_HCMVA PROTEIN UL87.	1.10e+01
12	53	56.4	231	10	VF16_VACCV PROTEIN F16.	1.64e+01
13	53	56.4	231	10	VF16_VACCC PROTEIN F16.	1.64e+01
14	53	56.4	725	11	YDC8_SCHPO HYPOTHETICAL 84.0 KD	1.64e+01
15	53	56.4	847	6	NIRB_ECOLI NITRITE REDUCTASE (NA	1.64e+01
16	53	56.4	1522	2	CINL_LOLBL SODIUM CHANNEL PROTEI	1.64e+01
17	53	56.4	155	8	REXX_PSEFL REGULATORY PROTEIN RE	2.42e+01
18	52	55.3	232	8	RFAY_SCHPO LIPOPOLYSACCHARIDE CO	2.42e+01
19	52	55.3	281	11	YD3B_SCHPO HYPOTHETICAL 33.1 KD	2.42e+01
20	52	55.3	330	8	QUTG_EWENI QUTG PROTEIN.	2.42e+01
21	52	55.3	468	11	YL34_CAEEL HYPOTHETICAL 53.6 KD	2.42e+01
22	52	55.3				

23	52	55.3	515	4	GLCM_MOUSE	GLUCOSYLKERAMIDASE PR	2.42e+01
24	52	55.3	551	11	YVD3_CAEEL	HYPOTHETICAL ZINC MET	2.42e+01
25	52	55.3	578	1	ATPA_METBA	ATP SYNTHASE ALPHA CH	2.42e+01
26	52	55.3	676	10	VI08_VACCV	PUTATIVE RNA HELICASE	2.42e+01
27	52	55.3	676	10	VI08_VARV	PUTATIVE RNA HELICASE	2.42e+01
28	52	55.3	676	10	VI08_VACCC	PUTATIVE RNA HELICASE	2.42e+01
29	52	55.3	702	1	ARYA_MANSE	ARYLPHORIN ALPHA SUBU	2.42e+01
30	52	55.3	1351	8	RPOM_YEAST	DNA-DIRECTED RNA POLY	2.42e+01
31	52	55.3	2161	8	RRPL_CDVO	RNA POLYMERASE BETA S	2.42e+01
32	51	54.3	311	3	F3ST_FLACH	FLAVONOL 3-SULFOTRANS	3.56e+01
33	51	54.3	350	11	YCXA_EUGGR	HYPOTHETICAL 40.9 KD	3.56e+01
34	51	54.3	371	2	CONG_BOVIN	CONGLUTININ PRECURSOR	3.56e+01
35	51	54.3	537	4	GAG_BAEVM	GAG POLYPROTEIN (CONT	3.56e+01
36	51	54.3	642	9	SYT_BUCAP	THREONYL-TRNA SYNTHET	3.56e+01
37	51	54.3	659	7	PMPT_HUMAN	70 KD PEROXISOMAL MEM	3.56e+01
38	51	54.3	659	7	PMPT_MOUSE	70 KD PEROXISOMAL MEM	3.56e+01
39	51	54.3	659	7	PMPT_RAT	70 KD PEROXISOMAL MEM	3.56e+01
40	51	54.3	703	1	ARYB_MANSE	ARYLPHORIN BETA SUBUN	3.56e+01
41	51	54.3	760	6	MEPA_MOUSE	MEPRIN A ALPHA-SUBUNI	3.56e+01
42	51	54.3	882	5	HSSI_RAT	HEPARAN SULFATE N-DEA	3.56e+01
43	51	54.3	882	5	HSSI_HUMAN	HEPARAN SULFATE N-DEA	3.56e+01
44	51	54.3	994	9	TR44_ECOLI	TRANSPOSASE FOR TRANS	3.56e+01
45	51	54.3	1023	9	TSCC_PSEAM	THIAZIDE-SENSITIVE SO	3.56e+01

## ALIGNMENTS

RESULT 1  
 ID VGLM\_BUNL7 STANDARD; PRT: 1441 AA.  
 AC P09612;  
 DT 01-MAR-1989 (REL. 10, CREATED)  
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
 DE M POLYPROTEIN PRECURSOR (CONTAINS: NONSTRUCTURAL PROTEIN NS-M;  
 DE GLYCOPROTEINS G1 AND G2).  
 GN M.  
 OS BUNYAVIRUS LA CROSSE (ISOLATE L74).  
 OC VIRIDAE: SS-RNA ENVELOPED VIRUSES; NEGATIVE-STRAND; BUNYAVIRIDAE;  
 OC BUNYAVIRUSES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 88089508.  
 RA GRADY L.J., SANDERS M.L., CAMPBELL W.P.;  
 RL J. GEN. VIROL. 68:3057-3071(1987).  
 RN [2]  
 RP SEQUENCE OF 1-46 FROM N.A.  
 RX MEDLINE; 82216937.  
 RA CLERX-VAN HAASTER C.M., AKASHI H., AUERIN D.D., BISHOP D.H.L.;  
 RL J. VIROL. 41:119-128(1982).  
 CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS INCLUDING NONSTRUCTURAL PROTEIN NS-M, GLYCOPROTEIN G1, AND GLYCOPROTEIN G2.  
 CC -1- SIMILARITY: TO OTHER BUNYAVIRUSES M SEGMENT PROTEINS.  
 CC EMBL; D00202; E19620; -  
 DR EMBL; D10370; G221047; -  
 DR EMBL; J02231; -; NOT\_ANNOTATED\_CDS.  
 DR PIR; A29377; GNVCUC.  
 KW POLYPROTEIN; GLYCOPROTEIN; TRANSMEMBRANE; NONSTRUCTURAL PROTEIN; SIGNAL.  
 KW SIGNAL.  
 FT CHAIN 1 13  
 FT CHAIN 14 1441 M POLYPROTEIN.  
 FT CHAIN 22 299 GLYCOPROTEIN G2.  
 FT CHAIN 300 473 NONSTRUCTURAL PROTEIN NS-M.  
 FT CHAIN 474 1441 GLYCOPROTEIN G1.  
 FT CARBOHYD 57 457 POTENTIAL.  
 FT CARBOHYD 245 245 POTENTIAL.  
 FT CARBOHYD 490 490 POTENTIAL.  
 FT CARBOHYD 1177 1177 POTENTIAL.  
 FT CONFLICT 40 40 C -> S (IN REF. 2).  
 SQ SEQUENCE 1441 AA; 162540 MW; 27C3FD6E CRC32;  
 Query Match 62.8%; Score 59; DB 10; Length 1441;  
 Best Local Similarity 55.6%; Pred. No. 1.41e+00;

Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 388 rflniaymy 396  
|||: |||  
Qy 3 RFLHSFTMY 11

RESULT 2  
ID YVASIYACCV STANDARD; PRT; 67 AA.  
AC P20528;  
DT 01-FEB-1991 (REL. 17, CREATED)  
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)  
DT 01-FEB-1993 (REL. 27, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 7.5 KD PROTEIN.  
GN A ORF 5 OR SALFC.  
OS VACCINIA VIRUS (STRAIN WR), AND VACCINIA VIRUS (STRAIN COPENHAGEN).  
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; POXVIRIDAE; CHORDOPOXVIRINAE;  
OR THORPOXVIRUSES.

[1]  
SEQUENCE FROM N.A.  
RC STRAIN-WR;  
RX MEDLINE; 91259063.  
RA SMITH G.L., CHAN Y.S., HOWARD S.T.;  
RN J. GEN. VIROL. 72:1349-1376(1991).  
[2]  
SEQUENCE FROM N.A.  
RC STRAIN-COPENHAGEN;  
RX MEDLINE; 91021027.  
RA GOEBEL S.J., JOHNSON G.P., PERKUS M.E., DAVIS S.W., WINSLOW J.P.,  
RA PAOLETTI E.;  
RN VIROLOGY 179:247-266(1990).  
[3]  
COMPLETE GENOME.  
RC STRAIN-COPENHAGEN;  
RA GOEBEL S.J., JOHNSON G.P., PERKUS M.E., DAVIS S.W., WINSLOW J.P.,  
RA PAOLETTI E.;  
RN VIROLOGY 179:517-563(1990).  
DR EMBL; M35027; G335532;  
DR PIR; F42525; F42525.  
DR PIR; JQ1824; JQ1824.  
KW HYPOTHETICAL PROTEIN.  
SQ SEQUENCE 67 AA; 7503 MW; 7884D470 CRC32;

Query Match 61.7%; Score 58; DB 11; Length 67;

Best Local Similarity 50.0%; Pred. No. 2.14e+00;  
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

55 vaflnafgmy 64  
|||: |||  
Qy 2 RFLHSFTMY 11

RESULT 3  
ID YP84CAEEL STANDARD; PRT; 790 AA.  
AC Q09225;  
DT 01-NOV-1995 (REL. 32, CREATED)  
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 90.9 KD PROTEIN C08B11.4 IN CHROMOSOME II.  
GN C08B11.4.  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; ACCELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.  
RN [1]  
SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA STIMMS M.;  
RN SUBMITTED (NOV-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; 246676; G576997;  
DR WORMPEP; C08B11.4; CE01474.  
KW HYPOTHETICAL PROTEIN.  
SQ SEQUENCE 790 AA; 90873 MW; 8B2A0D83 CRC32;

Query Match 61.7%; Score 58; DB 11; Length 790;

Best Local Similarity 50.0%; Pred. No. 2.14e+00;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 303 ikfilafsmymy 312  
|||: |||  
Qy 2 RFLHSFTMY 11

RESULT 4  
ID YRKL\_BACSU STANDARD; PRT; 174 AA.  
AC P54439;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL NAD(P)H OXIDOREDUCTASE IN BLTR-SPOIIC INTERGENIC REGION.  
GN YRKL.  
OS BACILLUS SUBTILIS.  
OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.  
RN [1]  
SEQUENCE FROM N.A.  
RC STRAIN-168 / JH642;  
RA KOBAYASHI Y., MIZUNO M., MASUDA S., TAKEMARU K., HOSONO S.,  
RA SATO T., TAKEUCHI M.;  
RN SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -1- SIMILARITY: STRONG, TO E. COLI YABF, YHER AND H. INFLUENZAE H11544.  
CC -1- SIMILARITY: STRONG, TO MAMMALIAN NAD(P)H DEHYDROGENASE (QUINONE).  
DR EMBL; D84432; G1303711;  
DR SUBTILIST; BG11778; YRKL.  
KW HYPOTHETICAL PROTEIN; OXIDOREDUCTASE.

SQ SEQUENCE 174 AA; 20093 MW; AD4BDCD8 CRC32;

Query Match 60.6%; Score 57; DB 11; Length 174;

Best Local Similarity 60.0%; Pred. No. 3.26e+00;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 140 mrflpafgmy 149  
|||: |||  
Qy 2 RFLHSFTMY 11

RESULT 5  
ID BGIA\_THEMEA STANDARD; PRT; 446 AA.  
AC Q08638;  
DT 01-OCT-1994 (REL. 30, CREATED)  
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)  
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
DE BETA-GLUCOSIDASE A (EC 3.2.1.21) (GENTIOBIASE) (CELLOBIASE) (BETA-D-GLUCOSIDE GLUCOHYDROLASE).  
GN BGIA.  
OS THERMOTOGA MARITIMA.  
OC PROKARYOTA; NOT YET CLASSIFIED.  
RN [1]  
SEQUENCE FROM N.A.  
RC STRAIN-MSE8 / DSM 3109;  
RX MEDLINE; 94104595.  
RA LIEBL W., GABELSBERGER J., SCHLEIFER K.H.;  
RN MOL. GEN. GENET. 242:111-115(1994).  
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-GLUCOSE RESIDUES WITH RELEASE OF BETA-D-GLUCOSE.  
CC -1- PATHWAY: CELLULOSE DEGRADATION.  
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF GLYCOSYL HYDROLASES.

DR EMBL; X74163; G393291;  
DR PIR; S34570; S34570.  
DR PIR; S41561; S41561.  
DR PROSITE; PS00572; GLYCOSYL HYDROL\_F1\_1.  
DR PROSITE; PS00572; GLYCOSYL HYDROL\_F1\_2.  
KW HYDROLASE; GLYCOSIDASE; CELLULOSE DEGRADATION.  
FT ACT\_SITE 166 186  
FT ACT\_SITE 351 351  
SQ SEQUENCE 446 AA; 51548 MW; 0690A48B CRC32;

Query Match 59.6%; Score 56; DB 1; Length 446;  
Best Local Similarity 50.0%; Pred. No. 4.92e+00;

Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 239 vrfmhqfny 248  
:||:| |  
QY 2 IRLHSFTMY 11

RESULT 6  
ID CRY5\_BACTA STANDARD; PRT; 1165 AA.  
AC P19415;  
DT 01-NOV-1990 (REL. 16, CREATED)  
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)  
DT 01-MAY-1991 (REL. 18, LAST ANNOTATION UPDATE)  
DE 132 KD CRYSTAL PROTEIN (DELTA ENDOTOXIN) (CRYSTALLINE ENTOMOCIDAL  
DE PROTOXIN).  
GN CRYD.  
OS BACILLUS THURINGIENSIS (SUBSP. AIZAWAI).  
OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
STRAIN=HD-68;  
MEDLINE; 91016842.  
HOEFTE H., SOETAERT P., JANSSENS S., PEPEROEN M.;  
NUCLEIC ACIDS RES. 18:5545-5545(1990).  
CC -!- FUNCTION: PROMOTES COLLOIDSMOTIC LYSIS BY BINDING TO THE MIDGUT  
CC EPIHELIAL CELLS OF MANY LEPIDOPTERAN LARVAE.  
CC -!- TOXIC SEGMENT OF THE PROTEIN IS LOCATED IN THE N-TERMINUS.  
CC -!- DEVELOPMENTAL STAGE: THE CRYSTAL PROTEIN IS PRODUCED DURING  
CC SPOULATION AND IS ACCUMULATED BOTH AS AN INCLUSION AND AS PART  
CC OF THE SPORE COAT.  
DR EMBL; X54160; G40280; -;  
DR EMBL; A27640; E199236; -;  
DR EMBL; A15537; G490266; -;  
DR HSSP; P07130; IDLC.  
KW TOXIN; SPOULATION.  
SQ SEQUENCE 1165 AA; 132481 MW; 0D99380C CRC32;

Query Match 59.6%; Score 56; DB 2; Length 1165;  
Best Local Similarity 60.0%; Pred. No. 4.92e+00;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 300 vdfhnsftiy 309  
:||:| |  
QY 2 IRLHSFTMY 11

RESULT 7  
ID MC5R\_HUMAN STANDARD; PRT; 325 AA.  
AC P3032;  
DT 01-OCT-1993 (REL. 27, CREATED)  
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)  
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
DE MELANOCORTIN-5 RECEPTOR (MC5-R) (MC-2).  
GN MC5R.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=PLACENTA;  
RX MEDLINE; 93384614.  
RA CHAJJANI V., MUCENIECE R., WIKBERG J.F.S.;  
RL BIOCHEM. BIOPHYS. RES. COMMUN. 195:866-873(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 94234987.  
RA GRIFFON N., MIGNON V., FACCHINETTI P., DIAZ J., SCHWARTZ J.C.,  
RA SOKOLOFF P.;  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=PLACENTA;  
RX MEDLINE; 95258173.

RA FATHI Z., IBEN L.G., PARKER E.M.;  
RL NEUROCHEM. RES. 20:107-113(1994).  
CC -!- FUNCTION: RECEPTOR FOR MSH (ALPHA, BETA AND GAMMA) AND ACTH. THE  
CC ACTIVITY OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE  
CC ADENYLATE CYCLASE. THIS RECEPTOR IS A POSSIBLE MEDIATOR OF THE  
CC IMMUNOMODULATION PROPERTIES OF MELANOCORTINS.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -!- TISSUE SPECIFICITY: IS EXPRESSED IN THE BRAIN BUT NOT IN THE  
CC MELANOMA CELLS.  
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
DR EMBL; Z25470; G397610; -;  
DR EMBL; L27080; G435600; -;  
DR EMBL; U08353; G522164; -;  
DR GCRDB; GCR\_0793; -;  
DR GCRDB; GCR\_0840; -;  
DR MIM; 600042; -;  
DR PROSITE; PS00237; G-PROTEIN\_RECEPTOR.  
KW G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;  
KW PHOSPHORYLATION; LIPOPROTEIN; PALMITATE.  
FT DOMAIN 1 37 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 38 61 1 (POTENTIAL).  
FT DOMAIN 62 73 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 74 97 2 (POTENTIAL).  
FT DOMAIN 98 114 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 115 138 3 (POTENTIAL).  
FT DOMAIN 139 155 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 156 179 4 (POTENTIAL).  
FT DOMAIN 180 186 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 187 211 5 (POTENTIAL).  
FT DOMAIN 212 239 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 240 265 6 (POTENTIAL).  
FT DOMAIN 266 273 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 274 297 7 (POTENTIAL).  
FT DOMAIN 298 325 CYTOPLASMIC (POTENTIAL).  
FT CARBOHYD 2 2 POTENTIAL.  
FT CARBOHYD 15 15 POTENTIAL.  
FT CARBOHYD 20 20 POTENTIAL.  
FT CARBOHYD 28 28 POTENTIAL.  
FT LIPID 311 311 PALMITATE (POTENTIAL).  
FT LIPID 312 312 PALMITATE (POTENTIAL).  
FT CONFLICT 149 149 R -> A (IN REF. 2).  
FT CONFLICT 221 234 ALPGASSARQRTSM -> LCPGPAALRGPAW (IN  
FT REF. 1).  
FT CONFLICT 297 297 F -> Y (IN REF. 2).  
SQ SEQUENCE 325 AA; 36600 MW; 237C436E CRC32;

Query Match 57.4%; Score 54; DB 6; Length 325;  
Best Local Similarity 55.6%; Pred. No. 1.10e+01;  
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 272 rfmshfny 280  
:||:| |  
QY 3 RFLHSFTMY 11

RESULT 8  
ID UL87\_HSV6U STANDARD; PRT; 772 AA.  
AC P24437;  
DT 01-MAR-1992 (REL. 21, CREATED)  
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE PROTEIN U58.  
GN U58 OR 5R.  
OS HERPES SIMPLEX VIRUS (TYPE 6 / STRAIN UGANDA-1102).  
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; BETAHERPESVIRINAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 90080132.  
RA LAWRENCE G.L., CHEE M., CRAXTON M.A., GOMPELS U.A., HONESS R.W.,  
RA BARRELL B.G.;  
RL J. VIROL. 64:287-299(1990).  
RN [2]  
RP SEQUENCE FROM N.A.

RX MEDLINE; 95266321.  
RA COMPELS U.A., NICHOLAS J., LAWRENCE G., JONES M., THOMSON B.J.,  
RA MARTIN M.E., EFSTATHIOU S., CRAXTON M., MACAULAY H.A.,  
RL VIROLOGY 209:29-51(1995).  
CC -!- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS TOGETHER EBV BCRFL,  
CC HSV-6 U58, HVS-1 24 AND HCMV UL87.  
DR EMBL; M68963; G325501; -.  
DR EMBL; X83413; G854037; -.  
DR PIR; F33560; F33560.  
SQ SEQUENCE 772 AA; 88746 MW; 9AD5BC1F CRC32;

Query Match 57.4%; Score 54; DB 10; Length 772;  
Best Local Similarity 85.7%; Pred. No. 1.10e+01;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 470 hsfvmyk 476  
||| |||  
6 HSFTMYK 12

RESULT 9  
ID UL87\_HSV7J STANDARD; PRT; 775 AA.  
AC P52363;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DE PROTEIN U58.  
GN U58.  
OS HERPES SIMPLEX VIRUS (TYPE 7 / STRAIN JT) (HHV7).  
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; BETAHERPESVIRINAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA NICHOLAS J.;  
RL SUBMITTED (JAN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -!- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS TOGETHER EBV BCRFL,  
CC HSV-6 U58, HVS-1 24 AND HCMV UL87.  
DR EMBL; U43400; G1139661; -.  
SQ SEQUENCE 775 AA; 89413 MW; C0C9D789 CRC32;

Query Match 57.4%; Score 54; DB 10; Length 775;  
Best Local Similarity 85.7%; Pred. No. 1.10e+01;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 461 hsfvmyk 467  
||| |||  
6 HSFTMYK 12

RESULT 10  
ID VNC5\_ADEEV STANDARD; PRT; 849 AA.  
AC P27454;  
DT 01-AUG-1992 (REL. 23, CREATED)  
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE NONCAPSID PROTEIN NS-1 (NONSTRUCTURAL PROTEIN NS1) (NCVP1).  
GN NS1.  
OS ADESI DENSOCYCLOSIS VIRUS (STRAIN GKV 002 002) (ADESI DENSOCYCLOSIS).  
OC VIRIDAE; SS-DNA NONENVELOPED VIRUSES; PARVOVIRIDAE; DENSOVIRUSES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 92024090.  
RA AFANASIEV B.N., GALLYOV E.E., BUCHATSKY L.P., KOZLOV Y.V.;  
RL VIROLOGY 185:323-336(1991).  
CC -!- FUNCTION: SEEMS NECESSARY FOR VIRAL DNA REPLICATION, PLAYS AN  
CC IS IMPORTANT ROLE IN THE REGULATION OF CAPSID GENE EXPRESSION. IT  
CC IS COVALENTLY BOUND TO THE 5' TERMINI OF VIRAL DNA, AND CAN  
CC INFLUENCE DNA PACKAGING. NS1 IS ALSO BELIEVED TO POSSESS THE  
CC TOPOISOMERASE I TYPE NICKING ACTIVITY, AND CYTOTOXIC ACTIVITY.  
DR EMBL; M37899; -; NOT\_ANNOTATED\_CDS.  
DR PIR; A40784; UYPAD.  
KW NONCAPSID PROTEIN; DNA REPLICATION; ATP-BINDING.  
FT NP\_BIND 624 631 ATP (POTENTIAL).  
SQ SEQUENCE 849 AA; 97543 MW; FFC7C288 CRC32;

Query Match 57.4%; Score 54; DB 10; Length 849;  
Best Local Similarity 30.0%; Pred. No. 1.10e+01;  
Matches 3; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Db 416 vkflrnyly 425  
:||||:|  
2 IRELSFTMY 11

RESULT 11  
ID UL87\_HCMVA STANDARD; PRT; 941 AA.  
AC P16730;  
DT 01-AUG-1990 (REL. 15, CREATED)  
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE PROTEIN UL87.  
GN UL87.  
OS HUMAN CYTOMEGALOVIRUS (STRAIN AD169).  
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; BETAHERPESVIRINAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 90269039.  
RA CHEE M.S., BANKIER A.T., BECK S., BOHNI R., BROWN C.M., CERNY R.,  
RA HORSNELL T., HUTCHISON C.A. III, KOUZARIDES T., MARTIGNETTI J.A.,  
RA PREDDIE E., SATCHELL S.C., TOWLINSON P., WESTON K.M., BARRELL B.G.;  
RL CURR. TOP. MICROBIOL. IMMUNOL. 154:125-169(1990).  
CC -!- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS TOGETHER EBV BCRFL,  
CC HSV-6 U58, HVS-1 24 AND HCMV UL87.  
DR EMBL; X17403; G59694; -.  
DR PIR; S09851; S09851.  
FT CARBOHYD 37 37 POTENTIAL.  
FT CARBOHYD 591 591 POTENTIAL.  
FT CARBOHYD 661 661 POTENTIAL.  
FT CARBOHYD 801 801 POTENTIAL.  
SQ SEQUENCE 941 AA; 104803 MW; A9C56CC3 CRC32;

Query Match 57.4%; Score 54; DB 10; Length 941;  
Best Local Similarity 85.7%; Pred. No. 1.10e+01;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 594 hsfvmyk 600  
||| |||  
6 HSFTMYK 12

RESULT 12  
ID VF16\_VARV STANDARD; PRT; 231 AA.  
AC P33874;  
DT 01-FEB-1994 (REL. 28, CREATED)  
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)  
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
DE PROTEIN F16.  
GN F16L OR C20L.  
OS VARIOLA VIRUS.  
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; POXVIRIDAE; CHORDOPOXVIRINAE.  
OC ORTHOPOXVIRUSES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN-INDIA-1967 / ISOLATE IND3;  
RX MEDLINE; 94152154.  
RA SHCHELKUNOV S.N., BLINOV V.M., RESENCHUK S.M., TOTMENIN A.V.,  
RA SANDAKHCHIEV L.S.;  
RL VIRUS RES. 30:239-258(1993).  
RN [2]  
RP COMPLETE GENOME.  
RP STRAIN-INDIA-1967 / ISOLATE IND3;  
RX MEDLINE; 93202281.  
RA SHCHELKUNOV S.N., BLINOV V.M., SANDAKHCHIEV L.S.;  
RL FEBS LETT. 319:80-83(1993).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX STRAIN-BANGLADESH-1975;

RX MEDLINE; 94088747.  
RA MASSUNG R.F., ESPOSITO J.J., LIU L., QI J., UTTERBACK T.R.,  
RA KNIGHT J.C., AUBIN L., YURAN T.E., PARSONS J.M., LOPAREV V.N.,  
RA SELIVANOV N.A., CAVALLARO K.F., KERLAVAGE A.R., MAHY B.W.J.,  
RA VENTER C.J.;  
RL NATURE 366:748-751(1993).  
DR EMBL; X69198; G297221; -.  
DR EMBL; L22579; G438958; -.  
DR PIR; B36841; B36841.  
SQ SEQUENCE 231 AA; 26497 MW; 096C212B CRC32;

Query Match 56.4%; Score 53; DB 10; Length 231;  
Best Local Similarity 50.0%; Pred. No. 1.64e+01;  
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 198 rfvnklkmyk 207  
|||: |||  
QY 3 RFLHSFTMYK 12

RESULT 13  
VF16\_VACCC STANDARD; PRT; 231 AA.  
P29892;  
DT 01-APR-1993 (REL. 25, CREATED)  
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)  
DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)  
DE PROTEIN F16.  
GN F16L OR F1.  
OS VACCINIA VIRUS (STRAIN L-1VP).  
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; POXVIRIDAE; CHORDOPOXVIRINAE;  
OC ORTHOPOXVIRUSES.  
RN [1]  
RA MIKRYUKOV N.N., CHIZHIKOV V.E., PRIKHOD'KO G.G., URMANOV I.M.,  
RA SERPINSKII O.I., BLINOV V.M., NIKOLIN A.E., VASILENKO S.K.;  
RL BIOTEKNOLOGIYA 4:442-449(1988).  
DR EMBL; M57977; G395696; -.  
SQ SEQUENCE 231 AA; 26550 MW; 32D97918 CRC32;

Query Match 56.4%; Score 53; DB 10; Length 231;  
Best Local Similarity 50.0%; Pred. No. 1.64e+01;  
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 198 rfvnklkmyk 207  
|||: |||  
QY 3 RFLHSFTMYK 12

RESULT 14  
VF16\_VACCC STANDARD; PRT; 231 AA.  
P21021;  
DT 01-FEB-1991 (REL. 17, CREATED)  
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)  
DT 01-FEB-1991 (REL. 17, LAST ANNOTATION UPDATE)  
DE PROTEIN F16.  
GN F16L.  
OS VACCINIA VIRUS (STRAIN COPENHAGEN).  
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; POXVIRIDAE; CHORDOPOXVIRINAE;  
OC ORTHOPOXVIRUSES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 91021027.  
RA GOEBEL S.J., JOHNSON G.P., PERKUS M.E., DAVIS S.W., WINSLOW J.P.,  
RA PAOLETTI E.;  
RL VIROLOGY 179:247-266(1990).  
RN [2]  
RP COMPLETE GENOME.  
RA GOEBEL S.J., JOHNSON G.P., PERKUS M.E., DAVIS S.W., WINSLOW J.P.,  
RA PAOLETTI E.;  
RL VIROLOGY 179:517-563(1990).  
DR EMBL; M35027; G335383; -.  
DR PIR; C42508; C42508.  
SQ SEQUENCE 231 AA; 26605 MW; C59DF26C CRC32;

Query Match 56.4%; Score 53; DB 10; Length 231;  
Best Local Similarity 50.0%; Pred. No. 1.64e+01;  
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

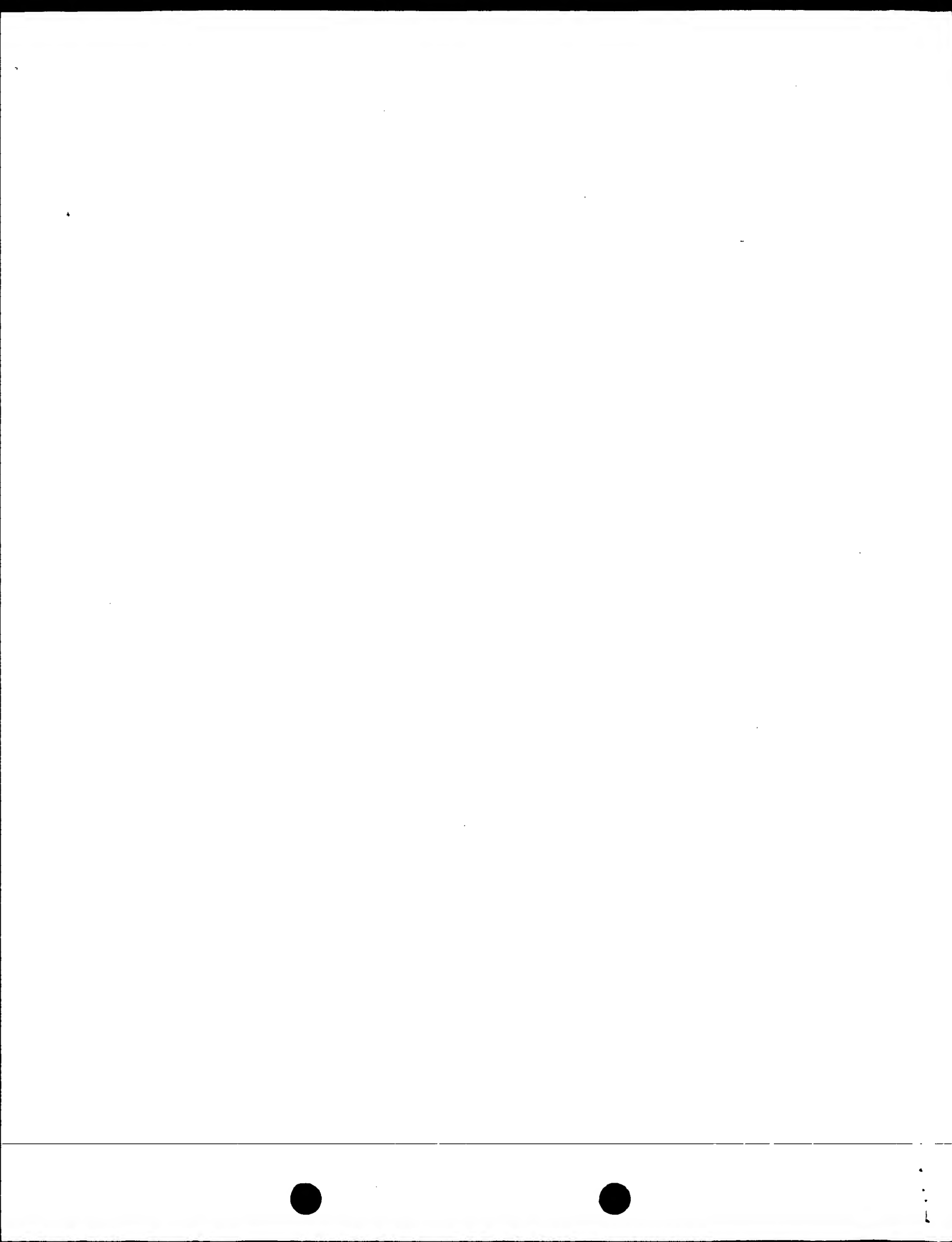
Db 198 rfvnklkmyk 207  
|||: |||  
QY 3 RFLHSFTMYK 12

RESULT 15  
ID YDC8\_SCHPO STANDARD; PRT; 725 AA.  
AC Q10425;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 84.0 KD PROTEIN C25G10.08 IN CHROMOSOME 1.  
GN SPAC25G10.08.  
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-972;  
RA MCLEAN J., HARRIS D., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;  
RL SUBMITTED (APR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -1- SIMILARITY: STRONG, TO YEAST PRT1/CDC63.  
DR EMBL; Z70691; E234481; -.  
KW HYPOTHETICAL PROTEIN.  
SQ SEQUENCE 725 AA; 84035 MW; 56EA1958 CRC32;

Query Match 56.4%; Score 53; DB 11; Length 725;  
Best Local Similarity 44.4%; Pred. No. 1.64e+01;  
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 229 rflhpyvkf 237  
|||: |||  
QY 3 RFLHSFTMY 11

Search completed: Tue Dec 2 14:45:47 1997  
Job time : 9 secs.



\*\*\*\*\*  
 M I S R L H  
 \*\*\*\*\*  
 (TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
 Distribution rights by IntelliGenetics, Inc.

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Dec 2 14:50:15 1997; MasPar time 8.93 Seconds  
 523.035 Million cell updates/sec

ular output not generated.

File: >US-08-915-004-4  
 Description: (1-380) from US08915004.pap  
 Perfect Score: 2861  
 Sequence: 1 ETFFPKYLHYDEETSHQLL.....QKLFLEMIGNQVSKISCL 380

Scoring table: PAM 150  
 Gap 11

Searched: 101610 seqs, 12294212 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: a-geneseq28  
 1:part1 2:part2 3:part3 4:part4 5:part6 6:part6 7:part7  
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
 14:part14 15:part15 16:part16 17:part17 18:part18  
 19:part19 20:part20 21:part21

Statistics: Mean 34.473; Variance 144.575; scale 0.238

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

#### SUMMARIES

ult No.	Score	Query Match	Length	ID	Description	Pred. No.
1	2861	100.0	380	20	Mature osteoclastogen	2.03e-275
2	2861	100.0	401	20	Full length osteoclas	2.03e-275
3	2847	99.5	401	20	Mutated OCIF, OCIF-C1	5.51e-274
4	2847	99.5	401	20	Mutated OCIF, OCIF-C2	5.51e-274
5	2843	99.4	401	20	Mutated OCIF, OCIF-C2	1.42e-273
6	2840	99.3	399	20	Mutated OCIF, OCIF-CL	2.87e-273
7	2841	99.3	401	20	Mutated OCIF, OCIF-C2	2.27e-273
8	2833	99.0	401	20	Mutated OCIF, OCIF-C2	1.50e-272
9	2794	97.7	393	20	Mutated OCIF, OCIF-DC	1.48e-268
10	2539	88.7	360	20	Mutated OCIF, OCIF-DC	1.89e-242
11	2337	88.7	390	20	Human tumour necrosis	3.03e-242
12	2481	86.7	351	20	Mutated OCIF, OCIF-CC	1.63e-236
13	2231	78.0	321	20	Mutated OCIF, OCIF-CS	5.85e-211
14	2218	77.5	359	20	Mutated OCIF, OCIF-DC	1.25e-209
15	2119	74.1	360	20	Mutated OCIF, OCIF-DC	1.60e-199
16	2079	72.7	359	20	Mutated OCIF, OCIF-DC	1.94e-195
17	1921	67.1	327	20	Mutated OCIF, OCIF-DD	2.52e-179
18	1915	66.9	272	20	Mutated OCIF, OCIF-CD	1.03e-178
19	1559	54.5	326	20	Mutated OCIF, OCIF-DD	1.72e-142
20	1370	47.9	197	20	Mutated OCIF, OCIF-CD	2.53e-123

21 1305 45.6 187 20 R99950 Mutated OCIF, OCIF-CB 9.58e-117  
 22 988 34.5 143 20 R99946 Mutated OCIF, OCIF-CC 8.69e-85  
 23 928 32.4 134 20 R99929 Osteoclastogenesis in 9.05e-79  
 24 924 32.3 145 20 R99930 Osteoclastogenesis in 2.28e-78  
 25 695 24.3 106 20 R99947 Mutated OCIF, OCIF-CC 1.52e-55  
 26 502 17.5 84 20 R99951 Mutated OCIF, OCIF-CP 1.43e-36  
 27 405 14.2 183 15 R77421 Bantp delta53 nerve g 3.43e-27  
 28 405 14.2 461 2 R11001 40KD TNF inhibitor pr 3.43e-27  
 29 401 14.0 461 13 R72504 p75 Tumour Necrosis F 8.28e-27  
 30 398 13.9 461 2 R11141 Human TNF-R deduced f 1.61e-26  
 31 398 13.9 461 8 R42058 Fibroblast derived TN 1.61e-26  
 32 398 13.9 485 2 R24016 Fusion protein TNFRFC 9.35e-26  
 33 390 13.6 518 9 R51003 Sequence of a recombi 2.53e-24  
 34 375 13.1 474 2 R11142 TNF-R deduced from mt 1.82e-23  
 35 366 12.8 461 9 R51002 Sequence of human tum 2.12e-19  
 36 323 11.3 392 2 R11605 Human 75KD TNF-bindin 2.43e-17  
 37 301 10.5 277 8 R38859 CD40 protein. 2.25e-14  
 38 269 9.4 326 5 R27866 Myxoma virus T2 prote 2.25e-14  
 39 269 9.4 326 15 R85072 Myxoma virus T2 prote 2.25e-14  
 40 260 9.1 325 15 R85071 Shope fibroma virus T 1.51e-13  
 41 260 9.1 325 15 R27865 Rabbit fibroma virus 1.51e-13  
 42 231 8.1 355 15 R85073 Cowpox virus T2-equiv 6.59e-11  
 43 223 7.8 283 21 W05809 Human tumour necrosis 3.46e-10  
 44 223 7.8 283 21 W12659 Human herpes simplex 3.46e-10  
 45 206 7.2 461 2 R07450 Rat Tumour Necrosis F 1.14e-08

#### ALIGNMENTS

RESULT 1  
 ID R99924 standard; Protein; 380 AA.  
 AC R99924:  
 DT 22-APR-1997 (first entry)  
 DE Mature osteoclastogenesis inhibitory factor.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 OS Homo sapiens.  
 PN W09626217-A1.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; J00374.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR N-PSDB: 136685.  
 PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 6; Page 62-64; 183pp: Japanese.  
 CC This sequence represents the mature osteoclastogenesis inhibitory  
 CC factor (OCIF) of the invention. The OCIF has a molecular weight by  
 CC SDS-PAGE of 60 kD under reducing conditions and 120 kD under non-  
 CC reducing conditions. The protein is adsorbed onto cation-exchangers  
 CC or heparin and its activity is lowered after 10 mins at 70 deg.C or  
 CC 30 mins at 56 deg.C, and is lost after 10 mins at 90 deg.C. OCIF is  
 CC useful in the control of bone resorption and therefore in the  
 CC treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 SQ Sequence 380 AA:

Query Match 100.0%; Score 2861; DB 20; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 2.03e-275;  
 Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 etfppkylhydeetshqlldckcpqgtylkchctakwtvcapcdpdytswhtsdecl 60  
 |||  
 QY 1 ETFFPKYLHYDEETSHQLLDCKCPPTGLKQCTAKWTVCAPCDPDHYTTSWHTSDECL 60  
 |||  
 Db 61 ycspsvckelqyvkqecnrthrvceckegryleiefclkhrcscppgfgvgaqgperntv 120  
 |||  
 QY 61 YCSPVCKELQYVKQECNRTNHRVCECKEGRYLEIEFCLKHRCSCPPGFGVGAQGTERTNTV 120  
 |||

Db 121 ckrpdpffsnetsskaperkhtncsvfllltqkgnathdnicsgnestqkcgidvltl 180  
 QY 121 CKRCPDGFSSNETSSKAPCRKHNCVSFGLLLTQKGNATHDNCSGNESTQKCGIDVTL 180  
 Db 181 ceaffrfavptkftpnwlsivdnlpgtkvnaesverikrhssgeqtfqllklwkhqn 240  
 QY 181 CEAFFRFAVPTKFTPNWLSIVDNLPGTKVNAESVERIKRHSSGEQTFQLLKLWKHQN 240  
 Db 241 kddiavkikiidilcensvqrhghantlfeqlrslmeslpgkxvgaediektikackp 300  
 QY 241 KDQDIVKKIIQDILCENSQVRHGHANTLFEQLRSLMESLPGKXVGAEDIEKTIKACKP 300  
 Db 301 sdqikllslwikngdgtlkgmlhalkhsktyhfpkvtqslkktirflhsfmykly 360  
 QY 301 SDQIKLLSLWIRKNGDQTLKGLMHALKHKSHTYHFPKVTQSLKKTIRFLHSFTMYKLY 360  
 Db 361 qklflemignqvskiscl 380  
 QY 361 QKLFLMIGNQVQSKISCL 380

## RESULT 2

ID R99925 standard; Protein; 401 AA.  
 AC R99925.  
 DT 22-APR-1997 (first entry)  
 DE Full length osteoclastogenesis inhibitory factor.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /note= "Signal peptide"  
 FT Protein 22..401  
 FT /note= "Mature OCIF, claim 6"  
 PN WO9626217-A1.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; J00374.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR N-PSDB: T36685.  
 PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PT Disclosure; Page 54-66; 183pp; Japanese.  
 CC This sequence represents the full length osteoclastogenesis inhibitory  
 CC factor (OCIF) of the invention. The OCIF has a molecular weight by  
 CC SDS-PAGE of 60 kD under reducing conditions and 120 kD under non-  
 CC reducing conditions. The protein is adsorbed onto cation-exchangers  
 CC or heparin and its activity is lowered after 10 mins at 70 deg.C or  
 CC 30 mins at 56 deg.C, and is lost after 10 mins at 90 deg.C. OCIF is  
 CC useful in the control of bone resorption and therefore in the  
 CC treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 SQ Sequence 401 AA;

Query Match 100.0%; Score 2861; DB 20; Length 401;  
 Best Local Similarity 100.0%; Pred. No. 2.03e-275;  
 Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 22 etfpkylhydeetshllcdkcpptgylkqhtakwtvcapcpdhytidswhstdecl 81  
 QY 1 ETFPPKYLHYDEETSHLLCDKCPPTGYLKQHTAKWTVCAPCPDHYTIDSWHSTDECL 60  
 Db 82 ycpvckelgyvqecnrthrvceckegryleiefclhrscppgfvvqagtperntv 141  
 QY 61 YCPVCKELGYVQECNRTNRVCECKEGRYLEIEFCLHRSCPPGFGVVQAGTPERNVT 120  
 Db 142 ckrpdpffsnetsskaperkhtncsvfllltqkgnathdnicsgnestqkcgidvltl 201  
 QY 121 CKRCPDGFSSNETSSKAPCRKHNCVSFGLLLTQKGNATHDNCSGNESTQKCGIDVTL 180

Db 202 ceaffrfavptkftpnwlsivdnlpgtkvnaesverikrhssgeqtfqllklwkhqn 261  
 QY 181 CEAFFRFAVPTKFTPNWLSIVDNLPGTKVNAESVERIKRHSSGEQTFQLLKLWKHQN 240  
 Db 262 kddiavkikiidilcensvqrhghantlfeqlrslmeslpgkxvgaediektikackp 321  
 QY 241 KDQDIVKKIIQDILCENSQVRHGHANTLFEQLRSLMESLPGKXVGAEDIEKTIKACKP 300  
 Db 322 sdqikllslwikngdgtlkgmlhalkhsktyhfpkvtqslkktirflhsfmykly 381  
 QY 301 SDQIKLLSLWIRKNGDQTLKGLMHALKHKSHTYHFPKVTQSLKKTIRFLHSFTMYKLY 360  
 Db 382 qklflemignqvskiscl 401  
 QY 361 QKLFLMIGNQVQSKISCL 380

## RESULT 3

ID R99931 standard; Protein; 401 AA.  
 AC R99931.  
 DT 22-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-C19S.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 OS Osteoporosis.  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /note= "Signal peptide"  
 FT Protein 22..401  
 FT /note= "Mature OCIF-C19S"  
 FT Misc difference 195  
 FT /label= C19S  
 PN WO9626217-A1.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; J00374.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR N-PSDB: T33161.  
 PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PT Claim 29; Page 94-96; 183pp; Japanese.  
 CC This sequence represents a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC sequence represents OCIF-C19S in which the 19th Cys residue in the  
 CC mature OCIF protein is substituted by Ser. The OCIF of the invention  
 CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
 CC and 120 kD under non-reducing conditions. The protein is adsorbed onto  
 CC cation-exchangers or heparin and its activity is lowered after 10 mins  
 CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
 CC deg.C. OCIF is useful in the control of bone resorption and therefore  
 CC in the treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 SQ Sequence 401 AA;

Query Match 99.5%; Score 2847; DB 20; Length 401;  
 Best Local Similarity 99.7%; Pred. No. 5.51e-274;  
 Matches 379; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 22 etfpkylhydeetshllcdkcpptgylkqhtakwtvcapcpdhytidswhstdecl 81  
 QY 1 ETFPPKYLHYDEETSHLLCDKCPPTGYLKQHTAKWTVCAPCPDHYTIDSWHSTDECL 60  
 Db 82 ycpvckelgyvqecnrthrvceckegryleiefclhrscppgfvvqagtperntv 141  
 QY 61 YCPVCKELGYVQECNRTNRVCECKEGRYLEIEFCLHRSCPPGFGVVQAGTPERNVT 120  
 Db 142 ckrpdpffsnetsskaperkhtncsvfllltqkgnathdnicsgnestqkcgidvltl 201  
 QY 142 CKRCPDGFSSNETSSKAPCRKHNCVSFGLLLTQKGNATHDNCSGNESTQKCGIDVTL 180



QY 121 CKRCPDGFSSNETSSKAPCRKHTNCVFGLLLTQKGNATHDNICSGNSESTQKCGIDVTL 180  
 Db 202 ceafffavptkftpnwlsvldnlpgtkvnaesverikrhssgeqtqllklwkhqn 261  
 QY 181 CEEAFFFAVPTKFTPNWLSVLDNLPGTKVNAESVERIKRQHSQSQTQFLLKLWKHQN 240  
 Db 262 kqdivvkiididlcensvqrhghantfeqlrslmeslpgkkgvgaediektikackp 321  
 QY 241 KQDIVVKIIDIIDLCENSQVQRHGHANTFEQURSLMESLPGRKKGVAEDIEKTIKACKP 300  
 Db 322 sdqilklslwrknqgdgtlkgimhalkhsktyhfpkvtqslkktirflhsftmykly 381  
 QY 301 SDQILKLLSLWRINKNGDQDTLKGMLHALKHKSITYHFPKVTQSLKKTIRFLHSFTWKLY 360  
 Db 382 qklflemignqvsvkiscsl 401  
 QY 361 OKLFLEMIGNQVSVKISCL 380

## RESULT 4

R9932 standard; Protein; 401 AA.  
 R9932;  
 22-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-C20S.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 KW osteoporosis.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /note= "Signal peptide"  
 FT Protein 22..401  
 FT /note= "Mature OCIF-C20S"  
 FT Misc\_difference 202  
 FT /label= C20S  
 FN WO9626217-A1.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; J00374.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 FA (SNOW) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR N-PSDB: T33162.  
 PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 32; Page 96-98; 183pp; Japanese.  
 CC This sequence represents a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC sequence represents OCIF-C20S in which the 20th Cys residue in the  
 CC mature OCIF protein is substituted by Ser. The OCIF of the invention  
 CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
 CC and 120 kD under non-reducing conditions. The protein is adsorbed onto  
 CC cation-exchangers or heparin and its activity is lowered after 10 mins  
 CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
 CC deg.C. OCIF is useful in the control of bone resorption and therefore  
 CC in the treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 SQ Sequence 401 AA;

Query Match 99.5%; Score 2847; DB 20; Length 401;  
 Best Local Similarity 99.7%; Pred. No. 5.51e-274;  
 Matches 379; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 22 etfppkylhydeetshqllcdkcpptgylkqhctakwtvcapcpdhyrdswhstdecl 81  
 QY 1 ETFPPKYLHYDEETSHQLLCDKCPPTGYLKQCTAKWKTVCAPCPDHYRDSWHTSDECL 60  
 Db 82 ycspvckelqyvkqecntrhvrceckegryleiefclkhrcscppgfgvvaqgperntv 141  
 QY 61 YCSVPCKELQYVKQECNTRHVRCECKEGRYLEIEFCLKHRSCPPGFGVVAQGPENTV 120  
 Db 142 ckrcpdgffsnetsskpcrkhtncsvfgllltqkgnathdnicsgnsesqkcgidvltl 201

QY 121 CKRCPDGFSSNETSSKAPCRKHTNCVFGLLLTQKGNATHDNICSGNSESTQKCGIDVTL 180  
 Db 202 ceafffavptkftpnwlsvldnlpgtkvnaesverikrhssgeqtqllklwkhqn 261  
 QY 181 CEEAFFFAVPTKFTPNWLSVLDNLPGTKVNAESVERIKRQHSQSQTQFLLKLWKHQN 240  
 Db 262 kqdivvkiididlcensvqrhghantfeqlrslmeslpgkkgvgaediektikackp 321  
 QY 241 KQDIVVKIIDIIDLCENSQVQRHGHANTFEQURSLMESLPGRKKGVAEDIEKTIKACKP 300  
 Db 322 sdqilklslwrknqgdgtlkgimhalkhsktyhfpkvtqslkktirflhsftmykly 381  
 QY 301 SDQILKLLSLWRINKNGDQDTLKGMLHALKHKSITYHFPKVTQSLKKTIRFLHSFTWKLY 360  
 Db 382 qklflemignqvsvkiscsl 401  
 QY 361 OKLFLEMIGNQVSVKISCL 380

## RESULT 5

R9933 standard; Protein; 401 AA.  
 AC R9933;  
 DT 22-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-C21S.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 KW osteoporosis.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /note= "Signal peptide"  
 FT Protein 22..401  
 FT /note= "Mature OCIF-C21S"  
 FT Misc\_difference 277  
 FT /label= C21S  
 FN WO9626217-A1.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; J00374.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR N-PSDB: T33163.  
 PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 35; Page 98-100; 183pp; Japanese.  
 CC This sequence represents a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC sequence represents OCIF-C21S in which the 21st Cys residue in the  
 CC mature OCIF protein is substituted by Ser. The OCIF of the invention  
 CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
 CC and 120 kD under non-reducing conditions. The protein is adsorbed onto  
 CC cation-exchangers or heparin and its activity is lowered after 10 mins  
 CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
 CC deg.C. OCIF is useful in the control of bone resorption and therefore  
 CC in the treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 SQ Sequence 401 AA;

Query Match 99.4%; Score 2843; DB 20; Length 401;  
 Best Local Similarity 99.2%; Pred. No. 1.42e-273;  
 Matches 377; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 22 etfppkylhydeetshqllcdkcpptgylkqhctakwtvcapcpdhyrdswhstdecl 81  
 QY 1 ETFPPKYLHYDEETSHQLLCDKCPPTGYLKQCTAKWKTVCAPCPDHYRDSWHTSDECL 60  
 Db 82 ycspvckelqyvkqecntrhvrceckegryleiefclkhrcscppgfgvvaqgperntv 141  
 QY 61 YCSVPCKELQYVKQECNTRHVRCECKEGRYLEIEFCLKHRSCPPGFGVVAQGPENTV 120

Db 142 ckrpdpffsnetsskpcrkhtncsvfllltqkgnathdncsgnsestqkgidvltl 201  
 QY 121 CKRCPDPFFSNETSSKAPCRKHTNCVSFGLLLTQKGNATHDNCSGNSESTQKCIDVTL 180  
 Db 202 ceeaffrfavptkftpnwslvldnlpgtknaesverikrhqssqeqtfgllklwkhqn 261  
 QY 181 CEEAFFRFAPVPTKFTPNWLSVLVDNLPGTKVNAESVERIKRHQSSQEQTFOLLKLWKHQN 240  
 Db 262 kqddvklkiqdidicnsqvrhghantfeqlrslmeslpgkkgvgaediektikackp 321  
 QY 241 KQDDIVKKLIQDIDICENSQVRHGHANTFEQLRSLMESLPGKKGVAEDIEKTIKACKP 300  
 Db 322 sdqilklslwikngdgtikglmhalkshktyhfpktvtqgslkktirflhtfmykly 381  
 QY 301 SDQILKLLSLWRIKNGDQDTLKLGMHALKSHKTYHFPKTVTQSLKKTIRFLHSFTMYKLY 360  
 Db 382 qklflemignqvskiscl 401  
 QY 361 QKLFLEMIGNQVQSVKISCL 380

## RESULT 6

ID R99942 standard; Protein; 399 AA.  
 AC R99942;  
 DT 23-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-CL.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /note= "Signal peptide"  
 FT Protein 22..399  
 FT /note= "Mature OCIF-CL"  
 PN WO9626217-A1.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; J00374.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 PI WPI: 96-402320/40.  
 DR N-PSDB: T33172.  
 PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 Claim 62; Page 117-119; 183pp; Japanese.  
 PS This sequence represents a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC mature OCIF protein is deleted. The OCIF of the invention  
 CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
 CC and 120 kD under non-reducing conditions. The protein is adsorbed onto  
 CC cation-exchangers or heparin and its activity is lowered after 10 mins  
 CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
 CC deg.C. OCIF is useful in the control of bone resorption and therefore  
 CC in the treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 SQ Sequence 399 AA;

Query Match 99.3%; Score 2840; DB 20; Length 399;  
 Best Local Similarity 100.0%; Pred. No. 2.87e-273;  
 Matches 378; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 22 etfppkylhydeetshqllcdkcpptgylkqhtakwtvcapcdpdytdswhtsdecl 81  
 QY 1 ETFPPKYLHYDEETSHQLLCKDKCPPTGYLKQHTAKWKTVCAPCDPDHYTDSNHTSDECL 60  
 Db 82 ycsppvckelqyvkgqecntrhnrvceckegryleiefclkhrcpppgfvgvqagtperntv 141  
 QY 61 YCSPVCKELQYVKQECNTRHNRYCECKEGRYLEIEFCLKHRSCPPPGFVGVAQTPERNTV 120  
 Db 142 ckrpdpffsnetsskpcrkhtncsvfllltqkgnathdncsgnsestqkgidvltl 201

QY 121 CKRCPDPFFSNETSSKAPCRKHTNCVSFGLLLTQKGNATHDNCSGNSESTQKCIDVTL 180  
 Db 202 ceeaffrfavptkftpnwslvldnlpgtknaesverikrhqssqeqtfgllklwkhqn 261  
 QY 181 CEEAFFRFAPVPTKFTPNWLSVLVDNLPGTKVNAESVERIKRHQSSQEQTFOLLKLWKHQN 240  
 Db 262 kqddvklkiqdidicnsqvrhghantfeqlrslmeslpgkkgvgaediektikackp 321  
 QY 241 KQDDIVKKLIQDIDICENSQVRHGHANTFEQLRSLMESLPGKKGVAEDIEKTIKACKP 300  
 Db 322 sdqilklslwikngdgtikglmhalkshktyhfpktvtqgslkktirflhsfmykly 381  
 QY 301 SDQILKLLSLWRIKNGDQDTLKLGMHALKSHKTYHFPKTVTQSLKKTIRFLHSFTMYKLY 360  
 Db 382 qklflemignqvskis 399  
 QY 361 QKLFLEMIGNQVQSVKIS 378

## RESULT 7

ID R99934 standard; Protein; 401 AA.  
 AC R99934;  
 DT 22-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-C22S.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /note= "Signal peptide"  
 FT Protein 22..401  
 FT /note= "Mature OCIF-C22S"  
 FT Misc difference 277  
 FT /label= C22S  
 PN WO9626217-A1.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; J00374.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 PI WPI: 96-402320/40.  
 DR N-PSDB: T33164.  
 PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 Claim 38; Page 100-102; 183pp; Japanese.  
 PS This sequence represents a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC mature OCIF protein is substituted by Ser. The OCIF of the invention  
 CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
 CC and 120 kD under non-reducing conditions. The protein is adsorbed onto  
 CC cation-exchangers or heparin and its activity is lowered after 10 mins  
 CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
 CC deg.C. OCIF is useful in the control of bone resorption and therefore  
 CC in the treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 SQ Sequence 401 AA;

Query Match 99.3%; Score 2841; DB 20; Length 401;  
 Best Local Similarity 99.5%; Pred. No. 2.27e-273;  
 Matches 378; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 22 etfppkylhydeetshqllcdkcpptgylkqhtakwtvcapcdpdytdswhtsdecl 81  
 QY 1 ETFPPKYLHYDEETSHQLLCKDKCPPTGYLKQHTAKWKTVCAPCDPDHYTDSNHTSDECL 60  
 Db 82 ycsppvckelqyvkgqecntrhnrvceckegryleiefclkhrcpppgfvgvqagtperntv 141  
 QY 61 YCSPVCKELQYVKQECNTRHNRYCECKEGRYLEIEFCLKHRSCPPPGFVGVAQTPERNTV 120

Db 142 ckrpdpffsnetsskpcrkhtncsvfllltqknathdnicsgnsestqkcgidvltl 201  
 QY 121 CKRCPDGFSSNETSSKAPCRKHTNCSVFGLLLTQKGNATHDNICSGNSESTQKCGIDVTL 180  
 Db 202 ceaffrfavptkftpnwlsvldnlpgtkvnaesverikrqhssqeqtfqlklwkhqn 261  
 QY 181 CEAFRFRAVPTKFTPNWLSVLDNLPGTKVNAESVERIKRQHSSEQEOTFQLKLWKHQN 240  
 Db 262 kqgdvkkliqdidlcensvqrhghanltfeqlrslmeslpgkkgvgaediektikackp 321  
 QY 241 KQGDIVKKLIQDIDLCENSVQRHGHANLTFEQLRSLMESUPGKKGVAEDIEKTIKACKP 300  
 Db 322 sqgilnlislrwknqgdgtlkglmhalkhsktyhfpktvtsqlkktirflhsftmykly 381  
 QY 301 SQGILNLSLRWKNGDQDTLKGMLHALKHSHKTYHFPKTVTQSLKKTIRFLHSFTMYKLY 360  
 Db 382 qklflemignqvsvkisc1 401  
 QY 361 QKLFLFLEMIGNQVSVKISCL 380

## ULT 8

R99935 standard; Protein; 401 AA.

AC R99935;  
 DT 22-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-C23S.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 OS Osteoporosis.  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /note= "Signal peptide"  
 FT Protein 22..401  
 FT /note= "Mature OCIF-C23S"  
 FT Misc.difference 400  
 FT /label= C23S  
 FN WO9626217-A1.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; J00374.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW ) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR N-PSDB: T33165.  
 PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 41; Page 103-105; 183pp; Japanese.  
 CC This sequence represents a mutated version of the full length  
 CC sequence represents OCIF-C23S in which the 23rd Cys residue in the  
 CC mature OCIF protein is substituted by Ser. The OCIF of the invention  
 CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
 CC and 120 kD under non-reducing conditions. The protein is adsorbed onto  
 CC cation-exchangers or heparin and its activity is lowered after 10 mins  
 CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
 CC deg.C. OCIF is useful in the control of bone resorption and therefore  
 CC in the treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 SQ Sequence 401 AA;

Query Match 99.0%; Score 2833; DB 20; Length 401;  
 Best Local Similarity 99.5%; Pred. No. 1.50e-272;  
 Matches 378; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 22 etfppkylhydeetsqllcdkcpptgtylkqhtakwtvcapcpdhyttdswhtsdesl 81  
 QY 1 ETFPFKYLHYDEETSQLLCDKCPPTGTYLKQHTAKWTVCAPCPDHYTTSWHTSDECL 60  
 Db 82 yscpvckelqyvkqecntrhnrvckegryleiefclkhrcppgfvagtperntv 141  
 QY 61 YCSPVCKELYVKQECNTRHNRVCKEGRYLEIEFCLKHRSCPPGFGVAGTPERNTV 120

Db 142 ckrpdpffsnetsskpcrkhtncsvfllltqknathdnicsgnsestqkcgidvltl 201  
 QY 121 CKRCPDGFSSNETSSKAPCRKHTNCSVFGLLLTQKGNATHDNICSGNSESTQKCGIDVTL 180  
 Db 202 ceaffrfavptkftpnwlsvldnlpgtkvnaesverikrqhssqeqtfqlklwkhqn 261  
 QY 181 CEAFRFRAVPTKFTPNWLSVLDNLPGTKVNAESVERIKRQHSSEQEOTFQLKLWKHQN 240  
 Db 262 kqgdvkkliqdidlcensvqrhghanltfeqlrslmeslpgkkgvgaediektikackp 321  
 QY 241 KQGDIVKKLIQDIDLCENSVQRHGHANLTFEQLRSLMESUPGKKGVAEDIEKTIKACKP 300  
 Db 322 sqgilnlislrwknqgdgtlkglmhalkhsktyhfpktvtsqlkktirflhsftmykly 381  
 QY 301 SQGILNLSLRWKNGDQDTLKGMLHALKHSHKTYHFPKTVTQSLKKTIRFLHSFTMYKLY 360  
 Db 382 qklflemignqvsvkisc1 401  
 QY 361 QKLFLFLEMIGNQVSVKISCL 380

## RESULT 9

ID R99948 standard; Protein; 393 AA.

AC R99948;  
 DT 23-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-CBst  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 OS Osteoporosis.  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /note= "Signal peptide"  
 FT Protein 22..393  
 FT /note= "Mature OCIF-CBst"  
 FT Misc.difference 392  
 FT /label= Gln371Leu  
 FN WO9626217-A1.  
 PD 29-AUG-1996; J00374.  
 PF 20-FEB-1996; JP-054977.  
 PR 20-FEB-1995; JP-207508.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW ) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR N-PSDB: T33178.

PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 80; Page 126-128; 183pp; Japanese.  
 CC This sequence represents a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC sequence represents OCIF-CBst in which Gln 371 is substituted by  
 CC Leu and amino acids 373-380 of the mature OCIF protein are deleted.  
 CC These changes are caused by the introduction of a restriction site in  
 CC the DNA encoding this protein. The OCIF of the invention has a  
 CC molecular weight by SDS-PAGE of 60 kD under reducing conditions  
 CC and 120 kD under non-reducing conditions. The protein is adsorbed onto  
 CC cation-exchangers or heparin and its activity is lowered after 10 mins  
 CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
 CC deg.C. OCIF is useful in the control of bone resorption and therefore  
 CC in the treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 SQ Sequence 393 AA;

Query Match 97.7%; Score 2794; DB 20; Length 393;  
 Best Local Similarity 99.7%; Pred. No. 1.48e-268;  
 Matches 371; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 22 etfppkylhydeetsqllcdkcpptgtylkqhtakwtvcapcpdhyttdswhtsdesl 81  
 QY 1 ETFPFKYLHYDEETSQLLCDKCPPTGTYLKQHTAKWTVCAPCPDHYTTSWHTSDECL 60

Db 82 ycspvckelqyvkqecnrthrvceckegryleiefclkhrcspgpgfvvgaqgperntv 141  
 QY 61 YCSPVCKELQYVKQECNRTHRVCECKEGRYLEIEFCLKHRCSPGPGFVGAQGPENTV 120  
 Db 142 chrpdgffsnetsskpcrkhtncsvfgllltkqgnathdnicsgnsesqtkcgidvltl 201  
 QY 121 CKRCPDGFSSNETSSKAPCRKHTNCSVFGLLTOKGNATHDNICSGNSESTQKCGIDVTL 180  
 Db 202 ceefrfaavtkftfnvslvldnlpdkvnaesverikrghsgeqtfqllklwkhn 261  
 QY 181 CEEAFRFAVTKFTFNVLNLPDKVNAESVERIKRQHSQEQTFQLLKLWKHN 240  
 Db 262 kqdvkikiididicnsvgrhghantfeqlrslmeslpgkkgvgaediektikackp 321  
 QY 241 KDQDVKKIIQDIDLCENSVQRHGHANTFEQLRSLMESLPGKKVGAEDIEKTIKACKP 300  
 Db 322 sdqilklslwikngdtdtlkghalkhskthfktvtqskkkirfihstfmykly 381  
 QY 301 SDQILKLSLWIKNGDQDTLKGHLALKHSHKTYHFKPTVTSLSKKTIRFLHSFTMYKLY 360  
 Db 382 qklflemignlv 393  
 QY 361 OKLFLEMIGNOV 372

## RESULT 10

ID R99336 standard; Protein; 360 AA.  
 AC R99336;  
 DT 23-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-DCR1  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 OS osteoporosis.  
 FH Synthetic.  
 FT Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /note= "Signal peptide"  
 FT protein 22..360  
 FT /note= "Mature OCIF-DCR1"  
 FT MiscIdifference 22..23  
 FT /note= "Position of deletion, delta 2-42"  
 PN WO9626217-A1.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; J00374.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW) SNOW BRAND MILK PROD CO LTD.  
 Nakagawa N, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 WPI: 96-402320/40.  
 DR N-PSDB; T33186.  
 PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 44; Page 105-107; 183pp; Japanese.  
 CC This sequence represents a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC sequence represents OCIF-DCR1 in which amino acids 2-42 of the  
 CC mature OCIF protein are deleted. The OCIF of the invention  
 CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
 CC and 120 kD under non-reducing conditions. The protein is adsorbed onto  
 CC cation-exchangers or heparin and its activity is lowered after 10 mins  
 CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins. at 90  
 CC deg.C. OCIF is useful in the control of bone resorption and therefore  
 CC in the treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 SQ Sequence 360 AA;

Query Match 88.7%; Score 2539; DB 20; Length 360;  
 Best Local Similarity 98.3%; Pred. No. 1.89e-242;  
 Matches 341; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

Db 15 sikwtq-epcpdhyytdswhtsdeclycspvckelqyvkqecnrthrvceckegryle 73  
 QY 34 TAKWTKVCAPCPDHYTDSWHTSDECLYCSPVCKELQYVKQECNRTHRVCECKEGRYLE 93

Db 74 iefclkhrcspgpgfvvgaqgperntvckrcpdgffsnetsskpcrkhtncsvfglllt 133  
 QY 94 IEFCLKHRCSPGPGFVGAQGPENTVCKRCPDGFFSNETSSKAPCRKHTNCSVFGLLT 153  
 Db 134 kqgnathdnicsgnsesqtkcgidvltlceefrfaavtkftfnvslvldnlpdkvna 193  
 QY 154 QKGNATHDNICSGNSESTQKCGIDVTLCEEAFRFAVTKFTFNVLNLPDKVNA 213  
 Db 194 esverikrghsgeqtfqllklwkhnkddivkklididicnsvgrhghantfeq 253  
 QY 214 ESVERIKRQHSQEQTFQLLKLWKHNKDDIVKKIIQDIDLCENSVQRHGHANTFEQ 273  
 Db 254 lrslneslpgkkgvgaediektikackpsdqilklslwikngdtdtlkghalkhskt 313  
 QY 274 LRSLSMESLPGKKVGAEDIEKTIKACKPSDQILKLSLWIKNGDQDTLKGHLALKHSHK 333  
 Db 314 yhfptvtqskkkirfihstfmyklyqklflemignovqskiscl 360  
 QY 334 YHFPTVTQSLAKKTIRFLHSFTMYKLYOKLFLEMIGNOVQSKISCL 380

## RESULT 11

ID R99357 standard; Protein; 390 AA.  
 AC R99357;  
 DT 05-MAY-1997 (first entry)  
 DE Human tumour necrosis factor receptor.  
 KW Tumour necrosis factor; TNF; receptor; TNF-beta; ligand; tumour;  
 KW differentiation; immune response; autoimmune disease; inflammation;  
 KW septic shock; graft-versus-host; apoptosis.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /label= sig\_peptide  
 PN WO9628546-A1.  
 PD 19-SEP-1996.  
 PF 15-MAR-1995; U03216.  
 PR 15-MAR-1995; WO-003216.  
 PR 29-MAR-1995; ZA-002587.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Fleischmann RD, Greene JM;  
 DR WPI: 96-433821/43.  
 DR N-PSDB; T35475.  
 PT New human tumour necrosis factor receptor - used to develop prods.  
 for treating e.g. tumours, infection, auto:immune disease, graft  
 PT rejection, cytotoxicity or inflammation  
 PS Claim 14; Fig 1; 59pp; English.  
 CC The receptor binds to TNF and in particular, TNF-beta.  
 CC of the receptor may be used for screening for antagonists and agonists  
 CC of the receptor and for ligands for the receptor. Such agonists may  
 CC be used to inhibit the growth of tumours, to stimulate cellular  
 CC differentiation, to mediate the immune response and anti-viral  
 CC response, to regulate growth and provide resistance to certain  
 CC infections. The antagonists may be used therapeutically, to treat  
 CC autoimmune diseases, inflammation, septic shock, to inhibit graft-  
 CC versus-host reactions, and to prevent apoptosis.  
 SQ Sequence 390 AA;

Query Match 88.7%; Score 2537; DB 20; Length 390;  
 Best Local Similarity 100.0%; Pred. No. 3.03e-242;  
 Matches 337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 22 eufpkylyhdeethshllcdkcpptylkqhctakwtvcapcpdhyytdswhtsdecl 81  
 QY 1 ETFPEKYLHYDEETSHQLLCKDPCPGTYLKQHTAKWTKVCAPCPDHYTDSWHTSDECL 60  
 Db 82 ycspvckelqyvkqecnrthrvceckegryleiefclkhrcspgpgfvvgaqgperntv 141  
 QY 61 YCSPVCKELQYVKQECNRTHRVCECKEGRYLEIEFCLKHRCSPGPGFVGAQGPENTV 120  
 Db 142 chrpdgffsnetsskpcrkhtncsvfgllltkqgnathdnicsgnsesqtkcgidvltl 201  
 QY 121 CKRCPDGFSSNETSSKAPCRKHTNCSVFGLLTOKGNATHDNICSGNSESTQKCGIDVTL 180

Db 202 ceaaffrfavptkftpnwslvldnlpctkynaesverikrqhssgeqtfqllklwkhqn 261  
 QY 181 CEEAFFRFVAVPTKFTPNWLSVLVDNLPGTKVNAESVERIKRQHSQSQTQFLLKLWKHQN 240  
 Db 262 kdqdivkklididlcensvqrhghlanltfeqlrslmeslpgkkgvgaediektikackp 321  
 QY 241 KDQDIVKKIIQIDILCENSQVRHGHANLTFEQLRSLMESLPGKKGVAEDIEKTIKACKP 300  
 Db 322 sdqilklslwrikngdgtlkglmhalkshktyhfp 358  
 QY 301 SDQILKLSLWRIKNGDQDTLKGMLHALKSHKTYHFP 337

RESULT 12  
 ID R99943 standard; Protein; 351 AA.  
 AC R99943;  
 DT 23-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-CC.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 OS Osteoporosis.  
 FH Synthetic.  
 FT Peptide 1..21  
 FT /note= "signal peptide"  
 FT Protein 22..351  
 FT /note= "Mature OCIF-CC"  
 PN WO9626217-Al.  
 PD 29-AUG-1996; J00374.  
 PF 20-FEB-1996; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW ) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yasuda H;  
 PI WPI: 96-402320/40.  
 DR N-PSDB: T33173.  
 DR DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 65; Page 119-121; 183pp; Japanese.  
 CC This sequence represents a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC sequence represents OCIF-CC in which amino acids 331-380 of the  
 CC mature OCIF protein are deleted. The OCIF of the invention  
 CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
 CC and 120 kD under non-reducing conditions. The protein is adsorbed onto  
 CC cation-exchangers or heparin and its activity is lowered after 10 mins  
 CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
 CC deg.C. OCIF is useful in the control of bone resorption and therefore  
 CC in the treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 SQ Sequence 351 AA;

Query Match 86.7%; Score 2481; DB 20; Length 351;  
 Best Local Similarity 100.0%; Pred. No. 1.63e-236;  
 Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 22 etfppkylhydeetshqllcdkcpptylkqctakwtvcapcdpdytidswhsdecl 81  
 QY 1 ETFPPKYLHYDEETSHQLLDCDKCPPTYLKQCTAKWTVCAPCDPDHYTIDSWHTSDECL 60  
 Db 82 ycspvckelqyvkgecnrthrvceckegryleiefclkhrcppgfgvvqagtperntv 141  
 QY 61 YCSPVCKELQYVKGCNRTNHRVCECKEGRYLEIEFCLKHRCPPGFGVVQAGTPERNVT 120  
 Db 142 ckrcpdgffsnetsskapcrkhtncsvfgllltqkgnathdnicsgsestqcgidvltl 201  
 QY 121 CKRCPDGFSSNETSSKAPCRKHTNCVSFGLLLTQGNATHDNICSGNSESTQCGIDVTL 180  
 Db 202 ceaaffrfavptkftpnwslvldnlpctkynaesverikrqhssgeqtfqllklwkhqn 261  
 QY 181 CEEAFFRFVAVPTKFTPNWLSVLVDNLPGTKVNAESVERIKRQHSQSQTQFLLKLWKHQN 240

Db 262 kdqdivkklididlcensvqrhghlanltfeqlrslmeslpgkkgvgaediektikackp 321  
 QY 241 KDQDIVKKIIQIDILCENSQVRHGHANLTFEQLRSLMESLPGKKGVAEDIEKTIKACKP 300  
 Db 322 sdqilklslwrikngdgtlkglmhalkshk 351  
 QY 301 SDQILKLSLWRIKNGDQDTLKGMLHALKH 330  
 RESULT 13  
 ID R99949 standard; Protein; 321 AA.  
 AC R99949;  
 DT 23-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-CSph.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 OS Osteoporosis.  
 FH Synthetic.  
 FT Key  
 FT Location/Qualifiers  
 FT Peptide 1..21  
 FT /note= "Signal peptide"  
 FT Protein 22..321  
 FT /note= "Mature OCIF-CSph"  
 PN WO9626217-Al.  
 PD 29-AUG-1996; J00374.  
 PF 20-FEB-1996; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW ) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yasuda H;  
 PI WPI: 96-402320/40.  
 DR N-PSDB: T33179.  
 DR DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 83; Page 128-129; 183pp; Japanese.  
 CC This sequence represents a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC sequence represents OCIF-CSph in which amino acids 298-380 of the mature  
 CC OCIF protein are replaced by Ser-Leu-Asp. These changes are caused by  
 CC the introduction of a restriction site in the DNA encoding this protein.  
 CC The OCIF of the invention has a molecular weight by SDS-PAGE of 60 kD  
 CC under reducing conditions and 120 kD under non-reducing conditions. The  
 CC protein is adsorbed onto cation-exchangers or heparin and its activity is  
 CC lowered after 10 mins at 70 deg.C or 30 mins at 56 deg.C, and is lost  
 CC after 10 mins at 90 deg.C. OCIF is useful in the control of bone  
 CC resorption and therefore in the treatment and prevention of disorders  
 CC of bone resorption, e.g. osteoporosis.  
 SQ Sequence 321 AA;

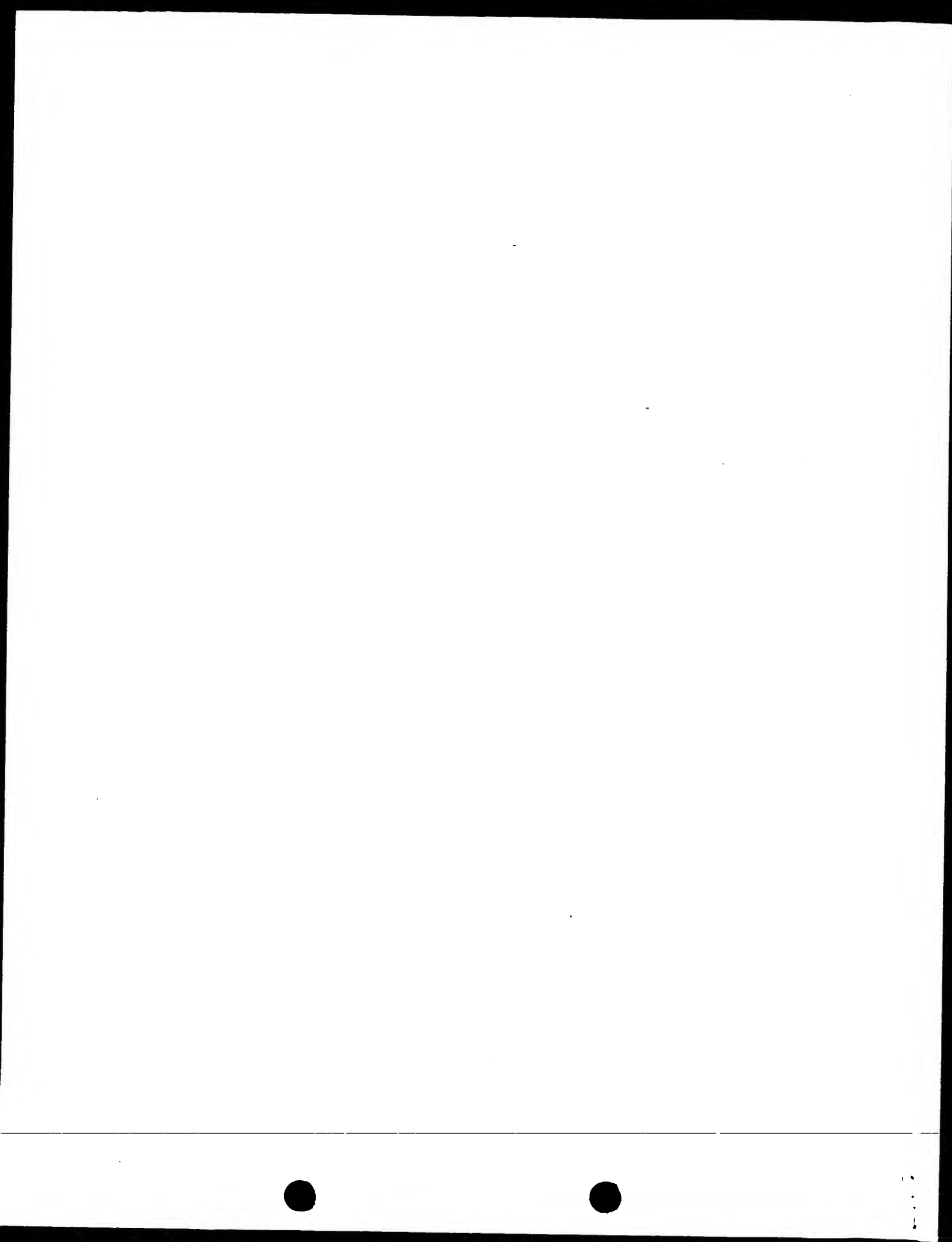
Query Match 78.0%; Score 2231; DB 20; Length 321;  
 Best Local Similarity 100.0%; Pred. No. 5.85e-211;  
 Matches 297; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 22 etfppkylhydeetshqllcdkcpptylkqctakwtvcapcdpdytidswhsdecl 81  
 QY 1 ETFPPKYLHYDEETSHQLLDCDKCPPTYLKQCTAKWTVCAPCDPDHYTIDSWHTSDECL 60  
 Db 82 ycspvckelqyvkgecnrthrvceckegryleiefclkhrcppgfgvvqagtperntv 141  
 QY 61 YCSPVCKELQYVKGCNRTNHRVCECKEGRYLEIEFCLKHRCPPGFGVVQAGTPERNVT 120  
 Db 142 ckrcpdgffsnetsskapcrkhtncsvfgllltqkgnathdnicsgsestqcgidvltl 201  
 QY 121 CKRCPDGFSSNETSSKAPCRKHTNCVSFGLLLTQGNATHDNICSGNSESTQCGIDVTL 180  
 Db 202 ceaaffrfavptkftpnwslvldnlpctkynaesverikrqhssgeqtfqllklwkhqn 261  
 QY 181 CEEAFFRFVAVPTKFTPNWLSVLVDNLPGTKVNAESVERIKRQHSQSQTQFLLKLWKHQN 240  
 Db 262 kdqdivkklididlcensvqrhghlanltfeqlrslmeslpgkkgvgaediektika 318  
 QY 241 KDQDIVKKIIQIDILCENSQVRHGHANLTFEQLRSLMESLPGKKGVAEDIEKTIKA 297



US-08-915-004-4.rag

Wed Dec 3 07:57:55 1997

Job time : 91 secs.





\*\*\*\*\*  
W P E R E H  
\*\*\*\*\* (TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution Rights by Intelligenetics, Inc.

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Dec 2 14:48:32 1997; MasPar time 13.77 Seconds  
797.024 Million cell updates/sec  
Similar output not generated.

Title: >US-08-915-004-4  
Description: (1-380) from US08915004.pep  
Perfect Score: 2861  
Sequence: 1 ETFPKYLHYDETSQLLC.....QKLFLEMIGNVQSVKISCL 380

Scoring table: PAM 150  
Gap 11

Searched: 91006 seqs, 28888923 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pir51  
1:ann1 2:ann2 3:ann3 4:ann4 5:unann1 6:unann2 7:unann3  
8:unann4 9:unann5 10:unann6 11:unann7 12:unann8  
13:unann9 14:unann10 15:unann16:unrev

Statistics: Mean 46.240; Variance 103.115; scale 0.448

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	398	13.9	461	6	A35356	tumor necrosis facto
2	377	13.2	459	14	I48854	gene murine tumour n
3	375	13.1	474	6	B38634	tumor necrosis facto
4	303	10.6	277	13	A60771	B-cell activation pr
5	294	10.3	289	14	A46515	B cell-associated su
6	294	10.3	305	14	A46476	CD40 - mouse
7	269	9.4	326	2	GVZML	T2 protein - myxoma
8	260	9.1	325	6	B43692	T2 protein - rabbit
9	260	9.1	435	13	I54182	tumor necrosis facto
10	230	8.0	138	16	S32385	gene G4R protein - v
11	230	8.0	349	8	D36858	G4R protein - variol
12	221	7.7	454	14	I57826	tumor necrosis facto
13	221	7.7	454	2	GQMS1	tumor necrosis facto
14	220	7.7	461	2	GQRT1	tumor necrosis facto
15	215	7.5	416	6	JGN006	nerve growth factor
16	213	7.4	427	2	GQHUN	nerve growth factor
17	207	7.2	425	6	A26431	nerve growth factor
18	186	6.5	461	14	JC4302	tumor necrosis facto
19	178	6.2	595	13	A42086	CD30 antigen precurs
20	172	6.0	455	2	GQHUT1	tumor necrosis facto
21	162	5.7	260	2	A46517	CD27 antigen precurs

22	159	5.6	256	14	B32393	T-cell antigen 4-lbB	4.60e-08
23	154	5.4	324	14	JC2395	Fas antigen - rat	2.33e-07
24	146	5.1	271	14	S12783	OX40 antigen precurs	3.01e-06
25	144	5.0	272	14	I48700	gene ox40 protein -	5.65e-06
26	141	4.9	255	13	JT0752	Lymphocyte activatio	1.44e-05
27	140	4.9	335	13	A38142	AP0-1 antigen, Fas a	1.97e-05
28	137	4.8	327	14	A46484	apoptosis-mediating	4.98e-05
29	134	4.7	250	2	A49053	CD27 antigen precurs	1.25e-04
30	134	4.7	314	13	I37383	FAS soluble protein	1.25e-04
31	134	4.7	335	13	A40036	apoptosis-mediating	1.25e-04
32	124	4.3	103	8	JQ1791	Salfl6R protein - va	2.48e-03
33	124	4.3	103	8	A42523	A53R protein - vacci	2.48e-03
34	115	4.0	360	11	S48365	hypothetical protein	3.31e-02
35	110	3.8	535	11	B34576	D2 protein precursor	1.33e-01
36	105	3.7	2813	3	VWUHU	von Willebrand facto	5.14e-01
37	103	3.6	344	11	S61037	hypothetical protein	8.74e-01
38	104	3.6	614	12	S43427	intermediate filamen	6.71e-01
39	102	3.6	3084	3	MMMSA	laminin chain A prec	1.14e+00
40	100	3.5	132	13	S57566	Fas/Apo-1/CD95 prote	1.91e+00
41	100	3.5	149	13	S58662	Fas-Delta-(4,7) prot	1.91e+00
42	101	3.5	713	11	JG6012	glutamine--fructose-	1.47e+00
43	101	3.5	1122	12	S64443	probable membrane pr	1.47e+00
44	100	3.5	2877	13	A38194	desmoplakin I - huma	1.91e+00
45	98	3.4	1947	3	S05697	myosin heavy chain C	3.19e+00

ALIGNMENTS

RESULT	1
ENTRY	A35356
TITLE	tumor necrosis factor receptor type 2 precursor - human
ALTERNATE_NAMES	75K tumor necrosis factor receptor
ORGANISM	#formal_name Homo sapiens #common_name man
DATE	14-Sep-1990 #sequence_revision 14-Sep-1990 #text_change 22-Nov-1996
ACCESSIONS	A35356; A38475; A48416; A36007; A23666; B35010; I38094
REFERENCE	A35356
#authors	Smith, C.A.; Davis, T.; Anderson, D.; Solam, L.; Beckmann, M.P.; Jerzy, R.; Dower, S.K.; Cosman, D.; Goodwin, R.G.
#journal	Science (1990) 248:1019-1023
#title	A receptor for tumor necrosis factor defines an unusual family of cellular and viral proteins.
#cross-references	MUID:90260639
#accession	A35356
##status	preliminary
##molecule_type	mRNA
##residues	1-461 #label SMI
##cross-references	GB:M32315
REFERENCE	A36475
#authors	Kohn, T.; Brewer, M.T.; Baker, S.L.; Schwartz, P.E.; King, M.W.; Hale, K.K.; Squires, C.H.; Thompson, R.C.; Vannice, J.L.
#journal	Proc. Natl. Acad. Sci. U.S.A. (1990) 87:8331-8335
#title	A second tumor necrosis factor receptor gene product can shed a naturally occurring tumor necrosis factor inhibitor.
#cross-references	MUID:91045991
#accession	A36475
##status	preliminary
##molecule_type	mRNA
##residues	1-195, 'R', 197-461 #label KOH
##cross-references	GB:M38549
REFERENCE	A48416
#authors	Dembic, Z.; Loetscher, H.; Gubler, U.; Pan, Y.C.; Lahm, H.W.; Gentz, R.; Brockhaus, M.; Lesslauer, W.
#journal	Cytokine (1990) 2:231-237
#title	Two human TNF receptors have similar extracellular, but distinct intracellular, domain sequences.
#cross-references	MUID:91370690
#accession	A48416
##status	preliminary
##molecule_type	mRNA; protein
##residues	23-461 #label DEM
##cross-references	NCBIN:63368; NCBIIP:63371



```

REFERENCE
#authors      A40254
               Goodwin, R.G.; Anderson, D.; Jerzy, R.; Davis, T.; Brannan,
               C.I.; Copeland, N.G.; Jenkins, N.A.; Smith, C.A.
#journal      Mol. Cell. Biol. (1991) 11:3020-3026
#title        Molecular cloning and expression of the type 1 and type 2
               murine receptors for tumor necrosis factor.
#cross-references MUID:91246168
#accession     A40254
               #molecule_type mRNA
               #residues 1-474 #label GOO
               #cross-references GB:M60469
REFERENCE
#authors      S54816
               Kissoneghis, M.; Fellowes, R.; Feldmann, M.; Chernaiovsky,
               Y.
#submission   submitted to the EMBL Data Library, May 1995
#description   Characterization of the promoter region of the murine p75-TNF
               receptor.
#accession     S54816
               #status preliminary
               #molecule_type DNA
               #residues 1-22 #label KIS
               #cross-references EMBL:X87128
ASSIGNATION   #superfamily tumor necrosis factor receptor type 2; NGF
               receptor repeat homology
FEATURE
23-474        #domain signal sequence #status predicted #label SIG\
               #product tumor necrosis factor receptor type 2 #status
               predicted #label MAT\
40-77         #domain NGF receptor repeat homology #label NG1\
79-120        #domain NGF receptor repeat homology #label NG2\
166-203       #domain NGF receptor repeat homology #label NG4
SUMMARY       #length 474 #molecular-weight 50319 #checksum 7767
               13.1%; Score 375; DB 6; Length 474;
Query Match   13.1%; Score 375; DB 6; Length 474;
Best Local Similarity 41.5%; Pred. No. 7.36e-43;
Matches 56; Conservative 21; Mismatches 61; Indels 11; Gaps 7;
Db 52 qmccakcpqgvykhfckntsdtvcaadceasmtygwnqfrtclscsscttdqveirac 111
   : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Qy 17 QLLCDKCPGTYLKQHTAKWTKVACPDHYTDSWHTSDECLYCSPVKELQVVKQEC 76
Db 112 tkqgnrvaceagrycalthsgscrqmrkscpgfgvassrapngnvckacapttf 171
   : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Qy 77 NRTNHRVCECKEGRY--LEIEF--CLKH-R-S-CPPGFGVQAGTPERTVCKRCPDGFF 129
   : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 172 sdttsdtdcrphricsi--laip--gnastdavcapes 206
   : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Qy 130 SNETSSKAPCRKHTNCSVFGLLLTQKGNATHDNCGNS 168
   : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :

RESULT 4
ENTRY        A60771 #type complete
TITLE        B-cell activation protein CD40 precursor - human
ALTERNATE_NAMES B-cell surface antigen Bp50
ORGANISM      #formal_name Homo sapiens #common_name man
DATE          03-Jun-1993 #sequence_revision 03-Feb-1994 #text_change
              06-Sep-1996
ACCESSIONS    S04460; A60771
REFERENCE      S04460
#authors      Stamenkovic, I.; Clark, E.A.; Seed, B.
#journal      EMBO J. (1989) 8:1403-1410
#title        A B-lymphocyte activation molecule related to the nerve
               growth factor receptor and induced by cytokines in
               carcinomas.
#cross-references MUID:89356608
#accession     S04460
               #molecule_type mRNA
               #residues 1-277 #label STA
               #cross-references EMBL:X60592
REFERENCE      A60771
#authors      Braesch-Andersen, S.; Paulie, S.; Koho, H.; Nika, H.;
               Aspenstroem, P.; Perlmann, P.
#journal      J. Immunol. (1989) 142:562-567

```

```

#title        Biochemical characteristics and partial amino acid sequence
               of the receptor-like human B cell and carcinoma antigen
               CDw40.
#accession     A60771
               #molecule_type protein
               #residues 21-50 #label BRA
               #experimental_source Burkitt lymphoma cell line Raj1
GENETICS
#gene          GDB:CD40
               #cross-references GDB:215268
               #map_position 20q12-20q13.2
KEYWORDS       B-cell; glycoprotein; phosphoprotein; transmembrane protein
FEATURE
1-20          #domain signal sequence #status predicted #label SIG\
21-277        #product B-cell activation protein CD40 #status
               experimental #label MAT\
21-193        #domain extracellular #status predicted #label EXT\
194-215       #domain transmembrane #status predicted #label TM\
216-277       #domain intracellular #status predicted #label CYT\
153,180       #binding_site carbohydrate (Asn) (covalent) #status
               predicted
SUMMARY       #length 277 #molecular-weight 30619 #checksum 6261
               10.6%; Score 303; DB 13; Length 277;
Query Match   10.6%; Score 303; DB 13; Length 277;
Best Local Similarity 36.8%; Pred. No. 1.19e-30;
Matches 56; Conservative 21; Mismatches 67; Indels 8; Gaps 7;
Db 38 csicqpgqkivscdtefctecicpgesefldtwrthchqhkycdpn-lglr-vqkg 95
   : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Qy 20 CDKCPGTYLKQHTAKWTKVACPDHYTDSWHTSDECL--YCPVKELQVVKQEC 76
   : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 96 tsetdtictceegwhctseacsvlhscspgfgvkiatgvsdticepcpvgffsnvs 155
   : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Qy 77 NRTNHRVCECKEGRY-L-EI-EFCLKHRSCPPGFGVQAGTPERTVCKRCPDGFFSNET 133
   : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 156 safekchpwtscetkdlvvqagtnktdvvcg 187
   : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Qy 134 SSKAPCRKHTNCSVFGLLLTQKGNATHDNCIS 165
   : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :

RESULT 5
ENTRY        A45515 #type complete
TITLE        B cell-associated surface molecule CD40 - mouse
ALTERNATE_NAMES #formal_name Mus musculus #common_name house mouse
ORGANISM      18-Jun-1993 #sequence_revision 18-Nov-1994 #text_change
DATE          03-Mar-1995
ACCESSIONS    A45515
REFERENCE      A45515
#authors      Grimaldi, J.C.; Torres, R.; Kozak, C.A.; Chang, R.; Clark,
               E.A.; Howard, M.; Cockayne, D.A.
#journal      J. Immunol. (1992) 149:3921-3926
#title        Genomic structure and chromosomal mapping of the murine CD40
               gene.
#cross-references MUID:93094586
#accession     A45515
               #status preliminary; not compared with conceptual translation
               #molecule_type nucleic acid
               #residues 1-289 #label GRI
               #cross-references NCBI:P120357
               #experimental_source BALB/c, liver
               #note sequence extracted from NCBI backbone
SUMMARY       #length 289 #molecular-weight 32111 #checksum 579
               10.3%; Score 294; DB 14; Length 289;
Query Match   10.3%; Score 294; DB 14; Length 289;
Best Local Similarity 38.8%; Pred. No. 3.73e-29;
Matches 59; Conservative 20; Mismatches 65; Indels 8; Gaps 6;
Db 38 cdicqpgsrtshtalektqchpcdsgefsagwnreirchqhrcpepn-qglr-vkkg 95
   : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Qy 20 CDKCPGTYLKQHTAKWTKVACPDHYTDSWHTSDECL--LY--CSPVKELQVVKQEC 76
   : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 96 taesdvtctckeghctskdceacaghtpctgfgvmmematettdtvcchpcpvgffsnvs 155
   : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :

```



```

##molecule_type mRNA
##residues 1-435 ##label RES
##cross-references GB:L04270; NID:q339761; CDS_PID:q339762
SUMMARY #length 435 #molecular-weight 46709 #checksum 63

Query Match 9.1%; Score 260; DB 13; Length 435;
Best Local Similarity 32.3%; Pred. No. 1.41e-23;
Matches 52; Conservative 23; Mismatches 75; Indels 11; Gaps 6;

Db 52 epqhricscppttyysakrsirdtvcatacaensynehwnylticqlcrpcdpv-mgl 110
Qy 13 ETSHQLLCKDCPPGTYLKQCHCTAKWKVCAPCPDHYTDSWHTSDEC-LY--CSPVCKEL 69
Db 111 eeiap-ctskrktcrqpgmfcaawalecthecellsdcpptgeaealkdvgkgnhcvp 169
Qy 70 QYVQECNRTHNRVCECKEGRY-----LETFCLKHRSCPPGFGV-VQAGTPERNTVCKR 123
Db 170 kcahfgntespsarcqhtrcengqlveaapgtagdttc 210
Qy 124 CPDGFSSNETSSKAPCRKHTNCSVFGLLLTQKGNATHDNIC 164

CULT 10
ENTRY S32385 #type fragment
TITLE gene G4R protein - variola virus (fragment)
ORGANISM #formal_name Variola virus
DATE 22-Nov-1993 #sequence_revision 22-Nov-1993 #text_change
ACCESSIONS S32385
REFERENCE S32385
#authors Shchelkunov, S.N.; Blinov, V.M.; Sandakhchiev, L.S.
#journal FEBS Lett. (1993) 319:80-83
#title Genes of variola and vaccinia viruses necessary to overcome
#accession S32385 the host protective mechanisms.
#status preliminary
#residues 1-138 ##label SHC
##cross-references EMBL:X69198
SUMMARY #length 138 #checksum 6036

Query Match 8.0%; Score 230; DB 16; Length 138;
Best Local Similarity 35.4%; Pred. No. 8.98e-19;
Matches 46; Conservative 19; Mismatches 58; Indels 7; Gaps 4;

Db 10 hnlclscppgttyasrlcdskntqctpgsgtftsrnnhlpaclscngcnsnqvtrts 69
Qy 16 HQLLCKDCPPGTYLKQCHCTAKWKVCAPCPDHYTDSWHTSDECILYSPVCKELQYVKQE 75
Db 70 cntthnricespggyellkgssgckacvsgtkcgigygvs-ghtsvgdvicspcgfgty 128
Qy 76 CNRTHNRVCECKEGRY--LE--IE--FCLKHRSCPPGFGVQAGTPERNTVCKRCPDGF 129
Db 129 shtvssadkc 138
Qy 130 SNETSSKAPC 139

RESULT 11
ENTRY D36858 #type complete
TITLE G4R protein - variola virus
ALTERNATE_NAMES B28R protein (COP)
ORGANISM #formal_name variola virus
DATE 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
ACCESSIONS D36858; S46888; S35987
REFERENCE A36859
#authors Blinov, V.M.
#submission submitted to GenBank, November 1992
#description not shown.
#accession D36858
#status preliminary
##molecule_type DNA
##residues 1-349 ##label BLI

```

```

##cross-references GB:X69198
##experimental_source strain India-1967, ssp. major, isolate Ind3
REFERENCE S46868
#authors Kolykhalov, A.A.; Blinov, V.M.; Gytarov, V.V.; Pozdnyakov,
S.G.; Chizhikov, V.E.; Frolov, I.V.; Totmenin, A.V.;
Shchelkunov, S.N.; Sandakhchiev, L.S.
#submission submitted to the EMBL Data Library, April 1992
#description Nucleotide sequence analysis of the region of variola virus
XhoI F O H P Q genome fragment.
#accession S46888
#status preliminary
##molecule_type DNA
##residues 1-349 ##label KOL
##cross-references EMBL:X67117
##experimental_source strain India-1967, isolate Ind3
CLASSIFICATION #superfamily NGF receptor repeat homology
FEATURE 68-109 #domain NGF receptor repeat homology #label NG2\
110-151 #domain NGF receptor repeat homology #label NG3
SUMMARY #length 349 #molecular-weight 38189 #checksum 2016

Query Match 8.0%; Score 230; DB 8; Length 349;
Best Local Similarity 35.4%; Pred. No. 8.98e-19;
Matches 46; Conservative 19; Mismatches 58; Indels 7; Gaps 4;

Db 40 hnlclscppgttyasrlcdskntqctpgsgtftsrnnhlpaclscngcnsnqvtrts 99
Qy 16 HQLLCKDCPPGTYLKQCHCTAKWKVCAPCPDHYTDSWHTSDECILYSPVCKELQYVKQE 75
Db 100 cntthnricespggyellkgssgckacvsgtkcgigygvs-ghtsvgdvicspcgfgty 158
Qy 76 CNRTHNRVCECKEGRY--LE--IE--FCLKHRSCPPGFGVQAGTPERNTVCKRCPDGF 129
Db 159 shtvssadkc 168
Qy 130 SNETSSKAPC 139

RESULT 12
ENTRY I57826 #type complete
TITLE tumor necrosis factor receptor - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change
ACCESSIONS I57826
REFERENCE I57826
#authors Rothe, J.G.; Bluethmann, H.; Gentz, R.; Lesslauer, W.;
Steinmetz, M.
#journal Mol. Immunol. (1993) 30:165-176
#title Genomic organization and promoter function of the murine
tumor necrosis factor receptor beta gene.
#cross-references MUID:93156721
#accession I57826
#status preliminary; translated from GB/EMBL/DBJ
##molecule_type DNA
##residues 1-454 ##label RES
##cross-references GB:M76656; NID:g202100; CDS_PID:g202102
GENETICS
#introns 13/3; 65/1; 108/1; 158/1; 184/2; 210/1; 248/1; 257/3; 353/1
#note gene name TNFR-2
SUMMARY #length 454 #molecular-weight 50030 #checksum 4267

Query Match 7.7%; Score 221; DB 14; Length 454;
Best Local Similarity 33.1%; Pred. No. 2.33e-17;
Matches 48; Conservative 21; Mismatches 65; Indels 11; Gaps 9;

Db 49 yvhsknnsicctkchkgtylvdsdpdpdrtdvcrecekgftasqnylrqlscsktrke 108
Qy 10 YDEETSHQLLCKDCPPGTYLKQCHCTAK-WKTVCAPCPDHYTDSWHTSDECILYSPVCKE 68
Db 109 msqveispqadktdvcgckengqfrylsethfcvdcspfcng-tvtipcktetntvcn 167
Qy 69 LOYVK-QECNRTHNRVCECKEG---RYL-EIEF-CLKHRSCPPGFGVQAGTPERNTVCK 122

```

```

Db 168 -chagflresecpvshckkneec 191
      ||| | : : | : | : |
QY 123 RCPDGFSSNETSSKA-P-CRKHTNC 145

RESULT 13
ENTRY QOMST1 #type complete
TITLE tumor necrosis factor receptor type 1 precursor - mouse
ALTERNATE_NAMES tumor necrosis factor receptor, 55K
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 18-Oct-1996
ACCESSIONS A38634; B40254; S16677; S19021; I54532
REFERENCE Lewis, M.; Tartaglia, L.A.; Lee, A.; Bennett, G.L.; Rice, G.C.; Wong, G.H.W.; Chen, E.Y.; Goeddel, D.V. Proc. Natl. Acad. Sci. U.S.A. (1991) 88:2830-2834
#journal Cloning and expression of cDNAs for two distinct murine tumor
#title necrosis factor receptors demonstrate one receptor is species specific.
#cross-references MUID:91187885
#accession A38634
#molecule_type mRNA
#residues 1-454 #label LEW
#cross-references GB:M60468
REFERENCE A40254
#authors Goodwin, R.G.; Anderson, D.; Jerzy, R.; Davis, T.; Brannan, C.I.; Copeland, N.G.; Jenkins, N.A.; Smith, C.A. Mol. Cell. Biol. (1991) 11:3020-3026
#journal Molecular cloning and expression of the type 1 and type 2
#title murine receptors for tumor necrosis factor.
#cross-references MUID:91246168
#accession B40254
#molecule_type mRNA
#residues 1-454 #label GO2
#cross-references GB:M60468
REFERENCE S16677
#authors Barrett, K.; Taylor-Fishwick, D.A.; Cope, A.P.; Kissonerghis, A.M.; Gray, P.W.; Feldmann, M.; Foxwell, B.M.J. Eur. J. Immunol. (1991) 21:1649-1656
#journal Cloning, expression and cross-linking analysis of the murine
#title p55 tumor necrosis factor receptor.
#cross-references MUID:91285014
#accession S16677
#molecule_type mRNA
#residues 1-454 #label BAR
#cross-references EMBL:X59238
REFERENCE S19021
#authors Rothe, J.G.; Brockhaus, M.; Gentz, R.; Lesslauer, W. Immunogenetics (1991) 34:338-340
#journal Molecular cloning and expression of the mouse Tnf receptor
#title type b.
#cross-references MUID:92039815
#accession S19021
#molecule_type mRNA
#residues 1-454 #label ROT
#cross-references EMBL:X57796
REFERENCE I54532
#authors Bebo, B.F. Immunogenetics (1994) 39:450-451
#journal Nucleotide sequence of the TNF type I receptor from a mouse
#title endotheliona cell line.
#cross-references MUID:94245292
#accession I54532
#status translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-454 #label RES
#cross-references GB:L26349; NID:g430732; CDS_PID:g430733
COMMENT This protein is one of two distantly related receptors for both TNF-alpha (cachectin) and TNF-beta (lymphotoxin).
CLASSIFICATION #superfamily tumor necrosis factor receptor type 1; NGF receptor repeat homology

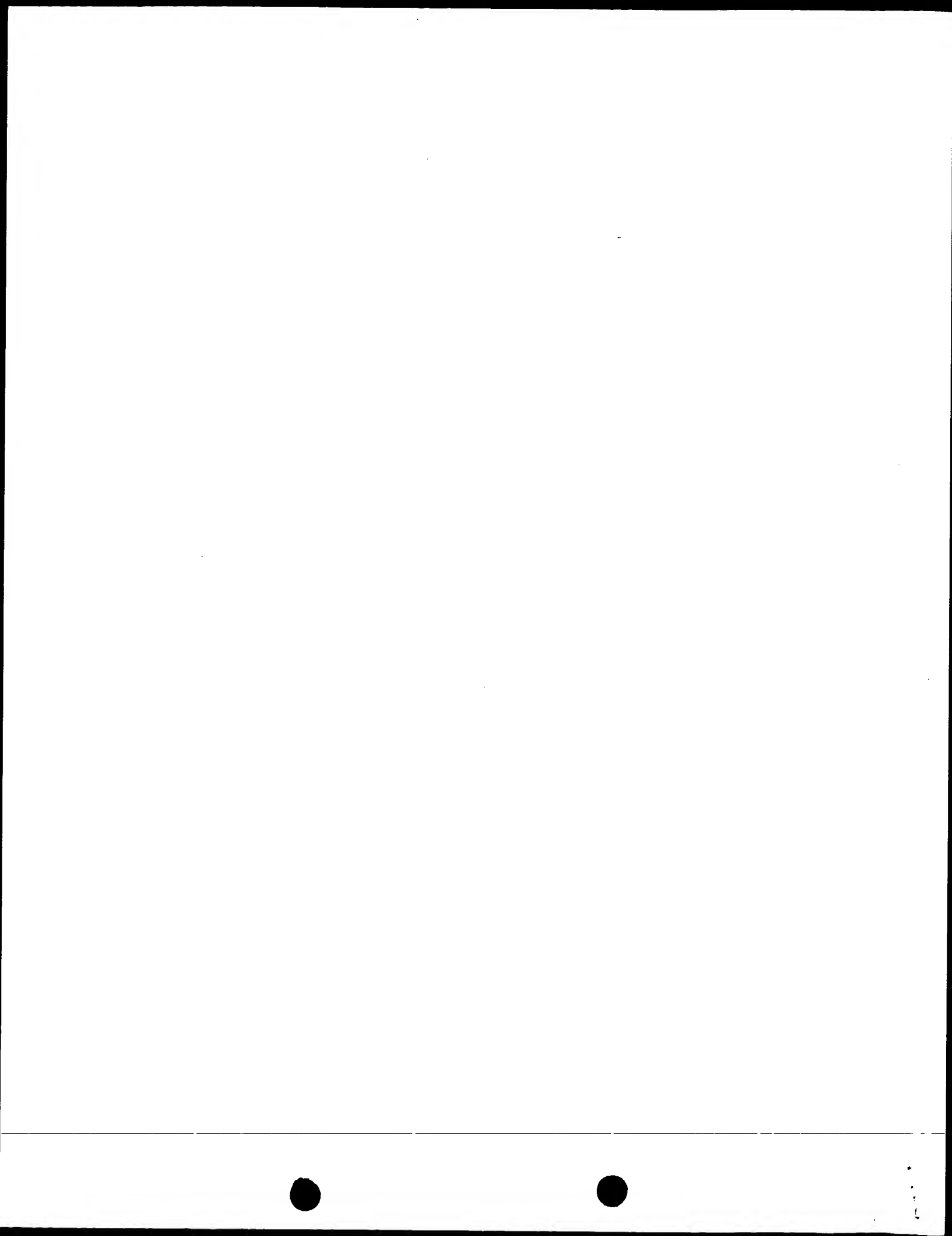
duplication; glycoprotein; receptor; transmembrane protein
#domain signal sequence #status predicted #label SIG\
#product tumor necrosis factor receptor type 1 #status predicted #label MAR\
#domain extracellular #status predicted #label EXT\
44-82 #domain NGF receptor repeat homology #label NG1\
84-126 #domain NGF receptor repeat homology #label NG2\
127-167 #domain NGF receptor repeat homology #label NG3\
168-204 #domain NGF receptor repeat homology #label NG4\
212-234 #domain transmembrane #status predicted #label MEM\
235-461 #domain intracellular #status predicted #label INT\
SUMMARY #length 454 #molecular-weight 50129 #checksum 4639

Query Match 7.7%; Score 221; DB 2; Length 454;
Best Local Similarity 33.11%; Pred. No. 2.33e-17;
Matches 48; Conservative 21; Mismatches 65; Indels 11; Gaps 9;
Db 49 yvshnnsicctkchkgtylvsdpcspgrdtvcrecekgftasqnylrgclscckrke 108
QY 10 YDEETSHQLLCDKCPGPGYLLKQHCTAK-WKIVCAPDPDHYTDSWHISDECLYCSPVCKE 68
Db 109 msqveispqadkdvogckengfrylsethfcqdcspofng-tvtipckettvnc 167
QY 69 LQYVK-QECNRTNHRVCEKEG---RYL-EIEF-CLKHRSOPPGFVVQAGTPERTVCK 122
Db 168 -chagflresecpvshckkneec 191
QY 123 RCPDGFSSNETSSKA-P-CRKHTNC 145

RESULT 14
ENTRY GQRTT1 #type complete
TITLE tumor necrosis factor binding protein 1 (TNF blocking factor)
CONTAINS #formal_name Rattus norvegicus #common_name Norway rat
ORGANISM 30-Jun-1992 #sequence_revision 07-Oct-1994 #text_change 05-Apr-1995
ACCESSIONS B36555
REFERENCE Himmler, A.; Maurer-Fogy, I.; Kroenke, M.; Scheurich, P.; Prizzenmaier, K.; Lantiz, M.; Olsson, I.; Hauptmann, R.; Stratowa, C.; Adolf, G.R. DNA Cell Biol. (1990) 9:705-715
#journal Molecular cloning and expression of human and rat tumor
#title necrosis factor receptor chain (p60) and its soluble derivative, tumor necrosis factor-binding protein.
#cross-references MUID:91090841
#accession B36555
#molecule_type mRNA
#residues 1-461 #label HIM
#cross-references GB:M63122
COMMENT This protein is one of two known receptors for both TNF-alpha (cachectin) and TNF-beta (lymphotoxin).
CLASSIFICATION #superfamily tumor necrosis factor receptor type 1; NGF receptor repeat homology
duplication; glycoprotein; receptor; transmembrane protein
#domain signal sequence #status predicted #label SIG\
#product tumor necrosis factor receptor type 1 #status predicted #label MAR\
#domain extracellular #status predicted #label EXT\
30-201 #product tumor necrosis factor binding protein #status predicted #label TBP\
44-82 #domain NGF receptor repeat homology #label NG1\
84-126 #domain NGF receptor repeat homology #label NG2\
127-167 #domain NGF receptor repeat homology #label NG3\
168-204 #domain NGF receptor repeat homology #label NG4\
212-234 #domain transmembrane #status predicted #label MEM\
235-461 #domain intracellular #status predicted #label INT\
#binding site carbohydrate (Asn) (covalent) #status predicted
SUMMARY #length 461 #molecular-weight 50969 #checksum 1617

```

Search completed: Tue Dec 2 14:49:58 1997  
Job time : 86 secs.





\*\*\*\*\*  
 M E S E L A  
 \*\*\*\*\*  
 (TM)

Release 2.1d John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
 Distribution rights by IntelliGenetics, Inc.

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
 Run on: Tue Dec 2 14:47:03 1997; MasPar time 9.74 Seconds  
 827.678 Million cell updates/sec  
 Similar output not generated.

Title: >US-08-915-004-4  
 Description: (1-380) from US08915004.pep  
 Perfect Score: 2861  
 Sequence: 1 ETFFPKYLHYDEETSHQLC.....QKLFLFMIGNQVQSVKISCL 380

Scoring table: PAM 150  
 Gap 11

Searched: 59021 segs, 21210388 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: swiss-prot34  
 1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
 8:part8 9:part9 10:part10 11:part11

Statistics: Mean 48.034; Variance 81.732; scale 0.588

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	405	14.2	461	9	TNR2_HUMAN TUMOR NECROSIS FACTOR	6.30e-63
2	375	13.1	474	9	TNR2_MOUSE TUMOR NECROSIS FACTOR	3.56e-56
3	303	10.6	277	2	CD40L_HUMAN CD40L RECEPTOR PRECUR	2.64e-40
4	294	10.3	289	2	CD40_MOUSE CD40L RECEPTOR PRECUR	2.31e-38
5	269	9.4	326	10	VT2_MXVL TUMOR NECROSIS FACTOR	5.03e-33
6	265	9.3	415	9	TNR2_MOUSE LYMPHOTOXIN-BETA RECE	3.52e-32
7	260	9.1	325	10	TNR2_SFVRA TUMOR NECROSIS FACTOR	3.98e-31
8	260	9.1	435	9	LYMPHOTOXIN-BETA RECE	3.98e-31
9	230	8.0	349	10	VC22_VARY PROTEIN C22/B28 HOMOL	6.59e-25
10	221	7.7	454	9	TNR1_MOUSE TUMOR NECROSIS FACTOR	4.45e-23
11	220	7.7	461	9	TNR1_RAT TUMOR NECROSIS FACTOR	7.09e-23
12	215	7.5	416	6	NGFR_CHICK LOW-AFFINITY NERVE GR	7.20e-22
13	213	7.4	427	6	NGFR_HUMAN LOW-AFFINITY NERVE GR	1.81e-21
14	207	7.2	425	6	NGFR_RAT LOW-AFFINITY NERVE GR	2.85e-20
15	188	6.6	323	3	FASA_BOVIN FASL RECEPTOR PRECURS	1.51e-16
16	186	6.5	461	9	TNR1_PIG TUMOR NECROSIS FACTOR	3.67e-16
17	178	6.2	595	2	CD30L_HUMAN CD30L RECEPTOR PRECUR	1.24e-14
18	172	6.0	455	9	TNR1_HUMAN TUMOR NECROSIS FACTOR	1.67e-13
19	159	5.6	256	1	41BB_MOUSE 4-1BB LIGAND RECEPTOR	4.22e-11
20	159	5.6	260	2	CD27L_HUMAN CD27L RECEPTOR PRECUR	4.22e-11
21	146	5.1	271	7	OX40_RAT OX40L RECEPTOR PRECUR	8.91e-09
22	144	5.0	272	7	OX40_MOUSE OX40L RECEPTOR PRECUR	2.00e-08

23	140	4.9	255	1	41BB_HUMAN 4-1BB LIGAND RECEPTOR	9.85e-08
24	141	4.9	277	7	OX40L_HUMAN OX40L RECEPTOR PRECUR	6.62e-08
25	137	4.8	327	3	FASA_MOUSE FASL RECEPTOR PRECUR	3.22e-07
26	134	4.7	250	2	CD27_MOUSE CD27L RECEPTOR PRECUR	1.04e-06
27	134	4.7	335	3	FASA_HUMAN FASL RECEPTOR PRECURS	1.04e-06
28	124	4.3	103	10	VA53_VACCC PROTEIN A53.	4.67e-05
29	124	4.3	103	10	VA53_VACCV PROTEIN A53.	4.67e-05
30	115	4.0	360	11	YIH9_YEAST HYPOTHETICAL 41.6 KD	1.24e-03
31	110	3.8	535	3	D2_DICDI CAMP-REGULATED D2 PRO	7.20e-03
32	105	3.7	2813	10	VWF_HUMAN VON WILLEBRAND FACTOR	3.96e-02
33	102	3.6	3084	6	LMAL_MOUSE LAMININ ALPHA-1 CHAIN	1.07e-01
34	101	3.5	712	4	GFAL_CANAL GLUCOSAMINE--FRUCTOSE	1.49e-01
35	101	3.5	1122	11	YG3C_YEAST HYPOTHETICAL 128.8 KD	1.49e-01
36	100	3.5	1752	3	DESP_HUMAN DESMOPLAKIN I AND II	2.05e-01
37	99	3.5	2670	11	YAO5_SCHPO PUTATIVE TRANSLATIONA	2.83e-01
38	96	3.4	431	7	PHOR_SHIDI PHOSPHATE REGULON SEN	7.33e-01
39	98	3.4	494	7	PR31_YEAST PRE-MRNA SPLICING FAC	3.90e-01
40	96	3.4	1104	9	SVY_YEAST VALYL-TRNA SYNTHETASE	7.33e-01
41	96	3.4	1203	10	KXPE_XENLA CHROMOSOME ASSEMBLY P	7.33e-01
42	96	3.4	1356	5	KAB7_YEAST PROBABLE SERINE/THREO	7.33e-01
43	98	3.4	1947	6	MYSC_CAEEL MYOSIN HEAVY CHAIN C	3.90e-01
44	96	3.4	3707	7	PGBM_MOUSE BASEMENT MEMBRANE-SPE	7.33e-01
45	97	3.4	4092	3	DYHC_YEAST DYNEIN HEAVY CHAIN, C	5.36e-01

## ALIGNMENTS

RESULT 1  
 ID TNR2\_HUMAN STANDARD; PRT: 461 AA.  
 AC P20333;  
 DT 01-FEB-1991 (REL. 17, CREATED)  
 DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)  
 DE TUMOR NECROSIS FACTOR RECEPTOR 2 PRECURSOR (TUMOR NECROSIS FACTOR  
 BINDING PROTEIN 2) (TBP1) (P80) (TNF-R2) (P75) (CD120B).  
 GN TNFR2 OR TNFR.  
 OS HOMO SAPIENS (HUMAN).  
 CC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 90260639.  
 RA SMITH C.A., DAVIS T., ANDERSON D., SOLAM L., BECKMANN M.P., JERZY R.,  
 RA DOWER S.K., COSMAN D., GOODWIN R.G.;  
 RL SCIENCE 248:1019-1023(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 9104591.  
 RA KOHNO T., BREWER M.T., BAKER S.L., SCHWARTZ P.E., KING M.W.,  
 RA HALE K.K., SQUIRES C.H., THOMPSON R.C., VANNICE J.L.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 87:8331-8335(1990).  
 RN [3]  
 RP SEQUENCE OF 116-461 FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE: 90349572.  
 RA HELLER R.A., SONG K., ONASCH M.A., FISCHER W.H., CHANG D.,  
 RA RINGOLD G.M.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 87:6151-6155(1990).  
 RN [4]  
 RP SEQUENCE OF 27-31.  
 RX MEDLINE: 90110215.  
 RA ENGELMANN H., NOVICK D., WALLACH D.;  
 RL J. BIOL. CHEM. 265:1531-1536(1990).  
 RN [5]  
 RP SEQUENCE OF 22-40; 65-69; 136-141; 300-306 AND 346-362.  
 RX MEDLINE: 91056048.  
 RA LOTSCHER H., SCHLAEGER E.J., LAHM H.-W., PAN Y.-C.E., LESSLAUER W.,  
 RA BROCKHAUS M.;  
 RL J. BIOL. CHEM. 265:20131-20138(1990).  
 RN [6]  
 RP CHARACTERIZATION.  
 RX MEDLINE: 93016040.  
 RA PENNICA D., LAM V.T., MIZE N.K., WEBER R.F., LEWIS M., FENDLY B.M.,  
 RA LIPARI M.T., GOEDEL D.V.;

RL J. BIOL. CHEM. 267:21172-21178(1992).  
 CC -1- FUNCTION: RECEPTOR FOR TNF-ALPHA. HIGH AFFINITY FOR TNA-ALPHA AND  
 CC APPROXIMATELY 5-FOLD LOWER AFFINITY FOR TNF-BETA.  
 CC -1- PTM: PHOSPHORYLATED: MAINLY ON SERINE RESIDUES WITH A VERY LOW  
 CC LEVEL ON THREONINE RESIDUES.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- SIMILARITY: CONTAINS A LA-NGFR/TNFR-TYPE CYSTEINE-RICH REGION.  
 DR EMBL; M32315; G189186; -;  
 DR EMBL; M35857; G339752; -;  
 DR EMBL; M55994; G339758; -;  
 DR PIR; A35356; A35356;  
 DR PIR; A36007; A36007;  
 DR PIR; A36475; A36475;  
 DR PIR; B35010; B35010;  
 DR PIR; A32666; A32666;  
 DR HSP; P19438; ITNR.  
 DR MIM; I91191; -;  
 DR PROSITE; PS00652; TNFR\_NGFR.  
 DR RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; REPEAT; SIGNAL;  
 PHOSPHORYLATION.  
 FT SIGNAL 1 22  
 FT CHAIN 23 461 TUMOR NECROSIS FACTOR RECEPTOR 2.  
 FT DOMAIN 23 257 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 258 287 POTENTIAL.  
 FT DOMAIN 288 461 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 39 201 4 X TNFR-CYS.  
 FT REPEAT 39 76 TNFR-CYS 1.  
 FT REPEAT 77 118 TNFR-CYS 2.  
 FT REPEAT 119 162 TNFR-CYS 3.  
 FT REPEAT 163 201 TNFR-CYS 4.  
 FT DISULFID 40 53 BY SIMILARITY.  
 FT DISULFID 54 67 BY SIMILARITY.  
 FT DISULFID 57 75 BY SIMILARITY.  
 FT DISULFID 78 93 BY SIMILARITY.  
 FT DISULFID 96 110 BY SIMILARITY.  
 FT DISULFID 100 118 BY SIMILARITY.  
 FT DISULFID 120 126 BY SIMILARITY.  
 FT DISULFID 134 143 BY SIMILARITY.  
 FT DISULFID 137 161 BY SIMILARITY.  
 FT DISULFID 164 179 BY SIMILARITY.  
 FT CARBOHYD 171 171 POTENTIAL.  
 FT CARBOHYD 193 193 POTENTIAL.  
 FT CONFLICT 141 141 R -> P (IN REF. 3).  
 FT CONFLICT 196 196 R -> M (IN REF. 1).  
 FT CONFLICT 363 363 A -> T (IN REF. 3).  
 SQ SEQUENCE 461 AA; 48316 MW; 0F5D0C44 CRC32;  
 Query Match 14.2%; Score 405; DB 9; Length 461;  
 Best Local Similarity 41.8%; Pred. No. 6.30e-63;  
 Matches 69; Conservative 26; Mismatches 59; Indels 11; Gaps 8;  
 Db 45 yvdqta-qmccskcspgghakvctktsdtycdscdedytqlwnwpeclscgscssd 103  
 QY 10 YDEETSHQLCDKCPGYLKHCHTAKWKVCAPCPDHYIDSWHTSDECLYCSPVKCEL 69  
 Db 104 qvetactreqnrictrpgycalskgegrclcaplkrpgrfgvarpgtetsdvckp 163  
 QY 70 QLVKQECNRTHNRVCEKRGY--LEI-EFC-L-KH-RSCPPGFGVQAGTPERTVCKR 123  
 Db 164 capgtfntsttdicrphqicnvvai--p-gnasrdavctsts 204  
 QY 124 CPDGFSSNETSKAPCRKHTNCVSFGLLLTKGNATHDNCSGNS 168  
 RESULT 2  
 ID TNR2\_MOUSE STANDARD; PRT; 474 AA.  
 AC P25119;  
 DT 01-MAY-1992 (REL. 22, CREATED)  
 DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE TUMOR NECROSIS FACTOR RECEPTOR 2 PRECURSOR (TNF-R2) (P75).  
 GN TNFR2 OR TNFR-2.  
 OS MUS MUSCULUS (MOUSE).

OC EUKARYOTA; METAFOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 91187885.  
 RA LEWIS M., TARTAGLIA L.A., LEE A., BENNETT G.L., RICE G.C.,  
 RA WONG G.H., CHEN E.X., GOEDDEL D.V.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 88:2830-2834(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 91246168.  
 RA GOODWIN R.G., ANDERSON D., JERZY R., DAVIS T., BRANNAN C.I.,  
 RA COPELAND N.G., JENKINS N.A., SMITH C.A.;  
 RL MOL. CELL. BIOL. 11:3020-3026(1991).  
 CC -1- FUNCTION: RECEPTOR FOR TNF-ALPHA.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- SIMILARITY: CONTAINS A LA-NGFR/TNFR-TYPE CYSTEINE-RICH REGION.  
 DR EMBL; M60469; G199828; -;  
 DR EMBL; M59378; G202095; -;  
 DR PIR; B38634; B38634;  
 DR HSP; P19438; ITNR.  
 DR PROSITE; PS00652; TNFR\_NGFR.  
 DR RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; REPEAT; SIGNAL.  
 FT SIGNAL 1 22  
 FT CHAIN 23 474 TUMOR NECROSIS FACTOR RECEPTOR 2.  
 FT DOMAIN 23 258 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 259 288 POTENTIAL.  
 FT DOMAIN 289 474 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 39 203 4 X TNFR-CYS.  
 FT REPEAT 39 77 TNFR-CYS 1.  
 FT REPEAT 78 119 TNFR-CYS 2.  
 FT REPEAT 120 164 TNFR-CYS 3.  
 FT REPEAT 165 203 TNFR-CYS 4.  
 FT DISULFID 40 54 BY SIMILARITY.  
 FT DISULFID 55 68 BY SIMILARITY.  
 FT DISULFID 58 76 BY SIMILARITY.  
 FT DISULFID 79 94 BY SIMILARITY.  
 FT DISULFID 97 111 BY SIMILARITY.  
 FT DISULFID 101 119 BY SIMILARITY.  
 FT DISULFID 121 127 BY SIMILARITY.  
 FT DISULFID 136 145 BY SIMILARITY.  
 FT DISULFID 139 163 BY SIMILARITY.  
 FT DISULFID 166 181 BY SIMILARITY.  
 FT CARBOHYD 69 69 POTENTIAL.  
 FT CARBOHYD 195 195 POTENTIAL.  
 SQ SEQUENCE 474 AA; 50319 MW; DC32B2B6 CRC32;  
 Query Match 13.1%; Score 375; DB 9; Length 474;  
 Best Local Similarity 41.5%; Pred. No. 3.56e-56;  
 Matches 66; Conservative 21; Mismatches 61; Indels 11; Gaps 7;  
 Db 52 qmccakcpvggqvkfhcnktsdtycdscdedytqlwnwpeclscgscssctdgvairac 111  
 QY 17 QLLCDKCPGTYLKHCHTAKWKVCAPCPDHYIDSWHTSDECLYCSPVKELQVVRQEC 76  
 Db 112 tkqqrvcacagrycalktsgscrcmqriskgpgfgvassrapngnvickacpgtf 171  
 QY 77 NTHRVCEKRGY--LEIEF--CLKH-R-S-CPPGFGVQAGTPERTVCKRCPDGF 129  
 Db 172 sdtstsdvcrphricsl--laip--gnasrdavcapes 206  
 QY 130 SNETSSKAPCRKHTNCVSFGLLLTKGNATHDNCSGNS 168  
 RESULT 3  
 ID CD40\_HUMAN STANDARD; PRT; 277 AA.  
 AC P25942;  
 DT 01-MAY-1992 (REL. 22, CREATED)  
 DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)  
 DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)  
 DE CD40L RECEPTOR PRECURSOR (B-CELL SURFACE ANTIGEN CD40) (BP50) (CDW40).  
 GN CD40.  
 OS HOMO SAPIENS (HUMAN).



[illegible]

RESULT	6	STANDARD	PRT	415 AA.
ID	TNRC_MOUSE			
AC	P50284;			
DT	01-OCT-1996	(REL. 34, CREATED)		
DT	01-OCT-1996	(REL. 34, LAST SEQUENCE UPDATE)		
DT	01-OCT-1996	(REL. 34, LAST ANNOTATION UPDATE)		
DE	LYMPHOXIN-BETA RECEPTOR PRECURSOR.			
DE	TNFCR.			
GN	MUS MUSCULUS (MOUSE).			
OS	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;			
OC	EUTHERIA; RODENTIA.			
OC	[1]			
RN	SEQUENCE FROM N.A.			
RP	STRAIN-CVB; TISSUE-LUNG;			
RC	FORCE W.R.; WILLIAMS-ABBOTT L.; BROWNING J.; HESSION C.; TIZARD R.;			
RA	WARE C.F.;			
RA	SUBMITTED (JUN-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.			
RL				

GENOMICS	30:312-319(1995).
-!	FUNCTION: RECEPTOR FOR THE LYMPHOTOXIN-BETA. POSSIBLE FUNCTION IN IMMUNE DEVELOPMENT.
-!	SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
-!	SIMILARITY: CONTAINS A LA-NGRF/TNFR-TYPE CYSTEINE-RICH REGION.
EMBL:	U29173; G881621;
EMBL:	L38423; G500223;
EMBL:	G30798; G1061327;
RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; REPEAT; SIGNAL. POTENTIAL.	
SIGNAL	1      30
CHAIN	31    415
DOMAIN	31    223
TRANSMEM	224   244
DOMAIN	245   415
DOMAIN	42    213
REPEAT	42    81
REPEAT	82    124
REPEAT	125   170
REPEAT	171   213
DISEULFD	43    58
DISEULFD	59    72
DISEULFD	62    80
DISEULFD	83    98
DISEULFD	101   116
DISEULFD	

FT	DISULFID	104	124	BY SIMILARITY.	
FT	DISULFID	126	132	BY SIMILARITY.	
FT	DISULFID	139	150	BY SIMILARITY.	
FT	DISULFID	142	169	BY SIMILARITY.	
FT	DISULFID	172	187	BY SIMILARITY.	
FT	CARBOHYD	40	40	POTENTIAL.	
FT	CARBOHYD	179	179	POTENTIAL.	
SEQ	SEQUENCE	415 AA; 44956 MW; 3C5DD121 CRC32;			

Query Match	9.3%;	Score 265;	DB 9;	Length 415;
Best Local Similarity	28.9%;	Pred. No. 3.52e-32;		
Matches	54;	Conservative 30;	Mismatches 92;	Indels 11; Gaps

Db	52	epmhdvccsrppggefavcvsrqdtvctkcpbnsynehwhlscqlrqp-cdiavlqf	110
QY	13	ETSHQLLCDRCPPGTYYKQHTAKTWVCAPCPDHYITDTSWHTSDCLYCSPPVCKE-LQY	71
Db	111	eevapctsdraecrcqpgmscvylndecvheeerlvlcqpgteaevtdlmdtdncv	170
QY	72	VK-QECNRTNRRVCEKGR---YLSEI--FCLKHR-S-CPPGFGV-VQAGTFERTNVC	122
Db	171	pkxpgfhqntssprarcqphtrceiagglveaapgtysdtkcnppepgamllal	230
QY	123	RCPDGFFSNETSSEKACRKHNTCSVFGLLLTQKGNATHDNICSGNSESTQKCGIDVTICE	182
Db	231	vlflft	237
QY	183	EAPFRFA	189

RESULT		7			
ID	VIT2_SFVKA	STANDARD;	PRT:	325 AA.	
AC	P25943;				
DT	01-MAY-1992	(REL. 22, CREATED)			
DT	01-MAY-1992	(REL. 22, LAST SEQUENCE UPDATE)			
DT	01-FEB-1994	(REL. 28, LAST ANNOTATION UPDATE)			
DE	TUMOR NECROSIS FACTOR SOLUBLE RECEPTOR PRECURSOR (PROTEIN T2).				
GN	T2.				
OS	SHOE FIBROMA VIRUS (STRAIN KASZA) (SFV).				
OC	VIRIDAE; DS-DNA ENVELOPED VIRUSES; POKVIRIDAE;				
ON	LEPORIPOVIRUSES.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE; 87321103.				
RA	UPTON C., DELANGE A.M., MCFADDEN G.;				
RL	VIROLOGY 160:20-30(1987).				
RN	[2]				
RP	FUNCTION.				
RX	MEDLINE; 91207415.				
RA	SMITH C.-A., DAVIS T., WIGNALL J.M., DIN W.S., FARRAH T., UPTON C.,				
RA	MCFADDEN G.G., GOODWIN R.G.;				
RL	BIOCHEM. BIOPHYS. RES. COMMUN. 176:335-342(1991).				
CC	-!- FUNCTION: BINDS TO TNF-ALPHA AND BETA. PROBABLY PREVENTS TNF TO				
CC	REACH CELLULAR TARGET AND THEREBY DEAMPENING THE POTENTIAL				
CC	ANTIVIRAL EFFECTS OF THE CYTOKINE.				
CC	-!- SIMILARITY: CONTAINS A LA-NGFR/TNFR-TYPE CYSTEINE-RICH REGION.				
DR	EMBL; M17433; -; NOT_ANNOTATED_CDS.				
DR	EMBL; A23727; E199408; -.				
DR	FIR; B43692; B43692.				
DR	HSP; P19438; ITNR.				
DR	PROSITE; PS00652; TNFR_NGFR.				
KW	RECEPTOR; GLYCOPROTEIN; REPEAT; SIGNAL.				
FT	SIGNAL	1	16	POTENTIAL.	
FT	CHAIN	17	325	PROTEIN T2.	
FT	DOMAIN	27	186	4 X TNFR-CYS.	
FT	REPEAT	27	62	TNFR-CYS 1.	
FT	REPEAT	63	104	TNFR-CYS 2.	
FT	REPEAT	105	147	TNFR-CYS 3.	
FT	REPEAT	148	186	TNFR-CYS 4.	
FT	CARBOHYD	105	105	POTENTIAL.	
FT	CARBOHYD	181	181	POTENTIAL.	
FT	CARBOHYD	205	205	POTENTIAL.	
FT	CARBOHYD	238	238	POTENTIAL.	







FT SIGNAL 1 28  
 FT CHAIN 29 427  
 FT DOMAIN 29 250  
 FT TRANSMEM 251 272  
 FT DOMAIN 273 427  
 FT DOMAIN 31 189  
 FT REPEAT 31 65  
 FT REPEAT 31 107  
 FT REPEAT 108 147  
 FT REPEAT 148 189  
 FT DOMAIN 344 421  
 FT DISULFID 32 43  
 FT DISULFID 44 57  
 FT DISULFID 47 64  
 FT DISULFID 67 83  
 FT DISULFID 86 99  
 FT DISULFID 89 107  
 FT DISULFID 109 122  
 FT DISULFID 125 138  
 FT DISULFID 128 146  
 FT DISULFID 149 164  
 FT DISULFID 167 180  
 FT DISULFID 170 188  
 FT DOMAIN 197 248  
 FT CARBOHYD 60  
 FT SEQUENCE 427 AA; 45183 MW; EE2924BD CRC32;

Query Match 7.4%; Score 213; DB 6; Length 427;  
 Best Local Similarity 34.9%; Pred. No. 1.81e-21;  
 Matches 45; Conservative 19; Mismatches 60; Indels 5; Gaps 5;

Db 44 ckacnlgegvagpcgan-qtvccepcldvtsfdvvsatepckpcte-cvlglsmsapcve 101  
 QY 20 CDKCPGTYLKHOKHTAKWKTCAPCDHY-YTDSWHTSDECLYSPVCKELQYVQKQCNR 78  
 Db 102 addavrcaygyvqdetgrceacrcvceagsglyfscqdkntvccepqdgtydsdeanhv 161  
 QY 79 THNRVCECKEGRYL-EI-EFCLKHRSCTPGGVQAGTPERTVCKRCPDGFSSNETSSK 136  
 Db 162 dpcclpctvc 170  
 QY 137 APCRKHTNC 145

RESULT 14  
 ID NGFR\_RAT STANDARD; PRT; 425 AA.  
 PO174;  
 01-APR-1988 (REL. 07, CREATED)  
 01-APR-1988 (REL. 07, LAST SEQUENCE UPDATE)  
 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 DE LOW-AFFINITY NERVE GROWTH FACTOR RECEPTOR PRECURSOR (NGF RECEPTOR)  
 DE (GF80)-LINGFR).  
 GN NGFR.  
 OS RATTUS NORVEGICUS (RAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 87115859.  
 RA RADEKE M.J., MISKO T.P., HSU C., HERZENBERG L.A., SHOOTER E.M.;  
 RL NATURE 325:593-597(1987).  
 RN [2]  
 RP SEQUENCE OF 1-22 FROM N.A.  
 RC TISSUE=LIVER.  
 RX MEDLINE; 93077038.  
 RA MEYSSIS M., TIMMUSK T., ALLIKMETS R., SAARMA M., PERSSON H.;  
 RL GENE 121:247-254(1992).  
 CC -!- FUNCTION: LOW AFFINITY RECEPTOR WHICH CAN BIND TO NGF, BDNF,  
 CC NT-3, AND NT-4.  
 CC -!- SUBUNIT: NGF RECEPTOR CAN FORM A HOMODIMER THROUGH DISULFIDE  
 CC BOND FORMATION.  
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -!- PTM: N- AND O-GLYCOSYLATED AND IS PHOSPHORYLATED ON SERINE.

CC -!- SIMILARITY: CONTAINS A LA-NGFR/TNFR-TYPE CYSTEINE-RICH REGION.  
 DR EMBL; X05137; G56756;  
 DR EMBL; X61269; -; NOT\_ANNOTATED\_CDS.  
 DR PIR; A26431; A26431.  
 DR HSSP; PI9438; ITNR.  
 DR PROSITE; PS00652; TNFR\_NGFR.  
 DR PROSITE; PS0017; DEATH\_DOMAIN.  
 KW RECEPTOR; NEUROGENESIS; TRANSMEMBRANE; GLYCOPROTEIN; REPEAT;  
 KW PHOSPHORYLATION; SIGNAL.  
 FT SIGNAL 1 29  
 FT CHAIN 30 425  
 FT DOMAIN 30 251  
 FT TRANSMEM 252 273  
 FT DOMAIN 274 425  
 FT DOMAIN 32 190  
 FT REPEAT 32 66  
 FT REPEAT 32 108  
 FT REPEAT 109 148  
 FT REPEAT 149 190  
 FT DOMAIN 198 249  
 FT DOMAIN 354 419  
 FT DISULFID 33 44  
 FT DISULFID 45 58  
 FT DISULFID 48 65  
 FT DISULFID 68 84  
 FT DISULFID 87 100  
 FT DISULFID 90 108  
 FT DISULFID 110 123  
 FT DISULFID 126 139  
 FT DISULFID 129 147  
 FT DISULFID 150 165  
 FT DISULFID 168 181  
 FT DISULFID 171 189  
 FT CARBOHYD 61  
 FT CARBOHYD 71  
 FT SEQUENCE 425 AA; 45432 MW; 7D78F258 CRC32;

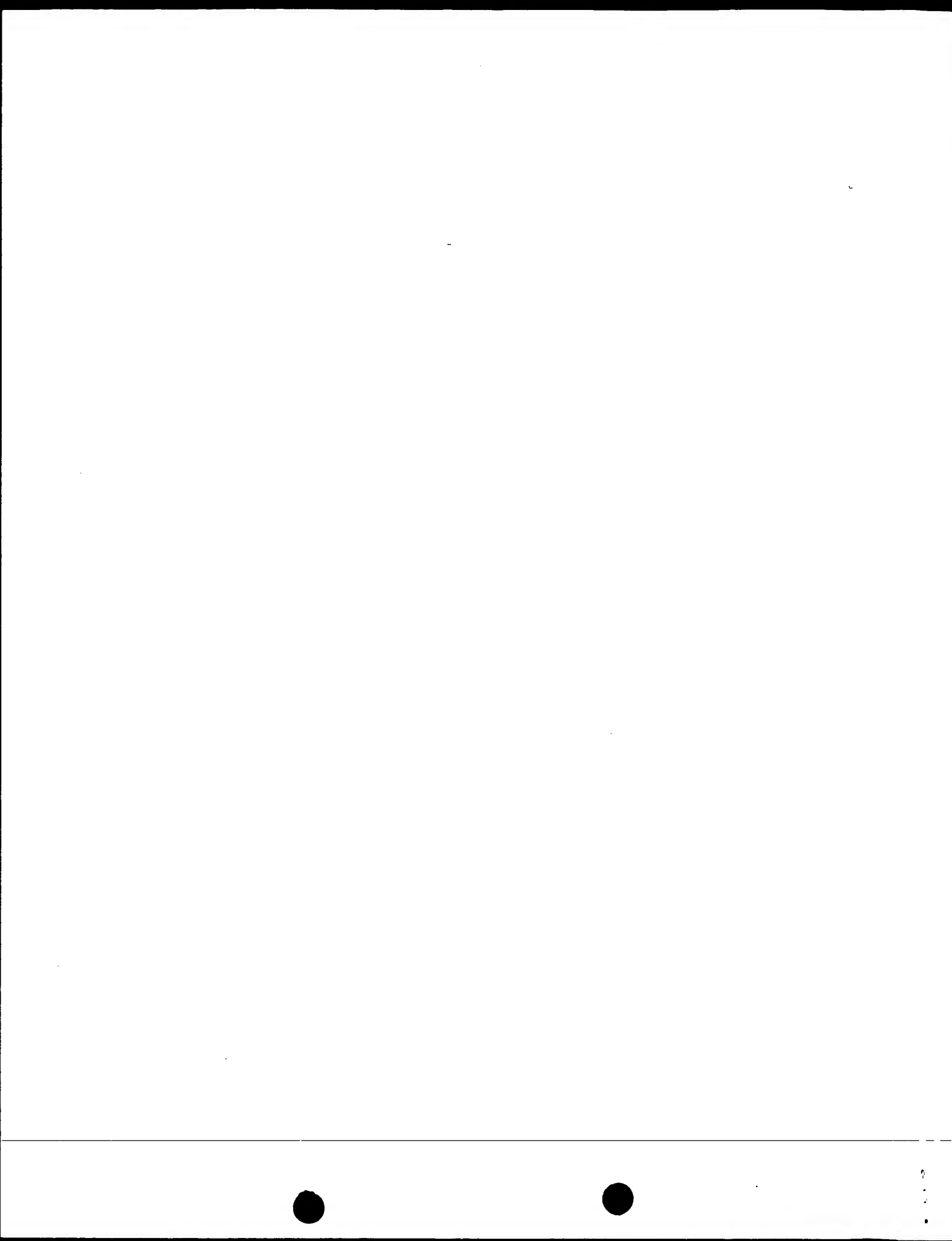
Query Match 7.2%; Score 207; DB 6; Length 425;  
 Best Local Similarity 33.3%; Pred. No. 2.85e-20;  
 Matches 43; Conservative 21; Mismatches 60; Indels 5; Gaps 4;

Db 45 ckacnlgegvagpcgan-qtvccepcldvtsfdvvsatepckpcte-cvlglsmsapcve 102  
 QY 20 CDKCPGTYLKHOKHTAKWKTCAPCDHY-YTDSWHTSDECLYSPVCKELQYVQKQCNR 78  
 Db 103 addavrcaygyvqdetgrceacrcvceagsglyfscqdkntvccepqdgtydsdeanhv 162  
 QY 79 THNRVCECKEGRYLEIE--FCLKHRSCTPGGVQAGTPERTVCKRCPDGFSSNETSSK 136  
 Db 163 dpcclpctvc 171  
 QY 137 APCRKHTNC 145

RESULT 15  
 ID FASA\_BOVIN STANDARD; PRT; 323 AA.  
 AC F51867;  
 DT 01-OCT-1996 (REL. 34, CREATED)  
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE FASL RECEPTOR PRECURSOR (APOPTOSIS-MEDIATING SURFACE ANTIGEN FAS)  
 DE (APO-1 ANTIGEN) (CD95).  
 GN APL1 OR FAS.  
 OS BOS TAURUS (BOVINE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; ARTIODACTYLA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 96226401.  
 RA YOO J., STONE R.T., BEATTIE C.W.;  
 RL DNA CELL BIOL. 15:227-234(1996).  
 CC -!- FUNCTION: RECEPTOR FOR A CYTOKINE LIGAND KNOWN AS FASL. MEDIATES  
 CC CELL DEATH. FAS-MEDIATED APOPTOSIS MAY HAVE A ROLE IN THE







\*\*\*\*\*

WIRE

(TM)

\*\*\*\*\*

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Dec 3 10:07:14 1997; Maspar time 2.12 Seconds  
86.963 Million cell updates/sec

ular output not generated.

File: >US-08-915-004-7  
Description: (1-15) from US08915004.pep  
Perfect Score: 113  
Sequence: 1 ETFPKYLHYDEETS 15

Scoring table: PAM 150  
Gap 15

Searched: 101610 seqs, 12294212 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq28  
1:part1 2:part3 3:part4 4:part5 5:part6 6:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21

Statistics: Mean 19.020; Variance 64.466; scale 0.295

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

sult No.	Score	Query Match	Length	ID	Description	Pred. No.
1	113	100.0	15	20	R99926 Osteoclastogenesis in	6.54e-05
2	113	100.0	84	20	R99951 Mutated OCIF, OCIF-CP	6.54e-05
3	113	100.0	106	20	R99947 Mutated OCIF, OCIF-CC	6.54e-05
4	113	100.0	143	20	R99946 Mutated OCIF, OCIF-CC	6.54e-05
5	113	100.0	145	20	R99930 Osteoclastogenesis in	6.54e-05
6	113	100.0	154	20	R99929 Osteoclastogenesis in	6.54e-05
7	113	100.0	187	20	R99950 Mutated OCIF, OCIF-CH	6.54e-05
8	113	100.0	197	20	R99945 Mutated OCIF, OCIF-CD	6.54e-05
9	113	100.0	272	20	R99944 Mutated OCIF, OCIF-CD	6.54e-05
10	113	100.0	321	20	R99949 Mutated OCIF, OCIF-CS	6.54e-05
11	113	100.0	326	20	R99940 Mutated OCIF, OCIF-DD	6.54e-05
12	113	100.0	327	20	R99941 Mutated OCIF, OCIF-DD	6.54e-05
13	113	100.0	351	20	R99943 Mutated OCIF, OCIF-CC	6.54e-05
14	113	100.0	359	20	R99937 Mutated OCIF, OCIF-DC	6.54e-05
15	113	100.0	359	20	R99939 Mutated OCIF, OCIF-DC	6.54e-05
16	113	100.0	360	20	R99938 Mature osteoclastogen	6.54e-05
17	113	100.0	380	20	R99924 Human tumour necrosis	6.54e-05
18	113	100.0	390	20	R99357 Mutated OCIF, OCIF-CL	6.54e-05
19	113	100.0	393	20	R99948 Mutated OCIF, OCIF-CL	6.54e-05
20	113	100.0	399	20	R99942 Mutated OCIF, OCIF-CL	6.54e-05

21	113	100.0	401	20	R99933 Mutated OCIF, OCIF-C2	6.54e-05
22	113	100.0	401	20	R99932 Mutated OCIF, OCIF-C2	6.54e-05
23	113	100.0	401	20	R99935 Mutated OCIF, OCIF-C2	6.54e-05
24	113	100.0	401	20	R99934 Mutated OCIF, OCIF-C1	6.54e-05
25	113	100.0	401	20	R99931 Mutated OCIF, OCIF-C1	6.54e-05
26	113	100.0	401	20	R99925 Full length osteoclas	6.54e-05
27	58	51.3	643	18	R97735 Bacillus thuringiensis	3.90e+01
28	58	51.3	643	1	P91462 67-KD protein toxin.	3.90e+01
29	56	49.6	330	5	R29643 PCTD ORF 8	6.05e+01
30	53	46.9	17	7	R37550 Platelet adhesion inh	1.16e+02
31	53	46.9	489	16	R88096 Thermostable beta-gal	1.16e+02
32	53	46.9	491	16	R88093 SDS-resistant hyperth	1.16e+02
33	53	46.9	491	16	R88095 Thermostable beta-gal	1.16e+02
34	52	46.0	107	16	R94385 Bettle patched protei	1.43e+02
35	52	46.0	147	14	R80521 Rat allograft inflamm	1.43e+02
36	52	46.0	223	3	R13495 P.denitrificans COB D	1.43e+02
37	52	46.0	700	15	R89145 Chondroitinase AC.	1.43e+02
38	51	45.1	343	4	R22463 Acetypolyamine amido	1.77e+02
39	51	45.1	464	17	R88069 Human ara Kb beta-gal	1.77e+02
40	51	45.1	1196	16	R75189 Osteoinductive retrov	1.77e+02
41	50	44.2	105	12	R67754 RNase T1	2.19e+02
42	50	44.2	540	12	R66547 Malolactic enzyme fro	2.19e+02
43	50	44.2	854	9	R44957 Feline Immunodeficien	2.19e+02
44	50	44.2	856	10	R51254 FIV DUTCH19K1 envelop	2.19e+02
45	50	44.2	870	9	R51246 Composite sequence of	2.19e+02

## ALIGNMENTS

## RESULT 1

ID R99926 standard; peptide; 15 AA.  
AC R99926;  
DT 22-APR-1997 (first entry)  
DE Osteoclastogenesis inhibitory factor N-terminal fragment.  
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
KW osteoporosis.  
OS Homo sapiens.  
PN WO9626217-A1.  
PD 29-AUG-1996.  
PF 20-FEB-1996; J00374.  
PR 20-FEB-1995; JP-054977.  
PR 21-JUL-1995; JP-207508.  
PA (SNOW) SNOW BRAND MILK PROD CO LTD.  
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
DR WPI: 96-402320/40.  
PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
PT for bone resorption control, esp. treatment of osteoporosis  
PS Claim 2; Page 67; 183pp; Japanese.  
CC The sequences given in R99921-23 and R99926 represent fragments of  
CC the osteoclastogenesis inhibitory factor (OCIF) of the invention.  
CC The OCIF has a molecular weight by SDS-PAGE of 60 kD under reducing  
CC conditions and 120 kD under non-reducing conditions. The protein is  
CC adsorbed onto cation-exchangers or heparin and its activity is  
CC lowered after 10 mins at 70 deg.C or 30 mins at 56 deg.C, and is  
CC lost after 10 mins at 90 deg.C. OCIF is useful in the control of  
CC bone resorption and therefore in the treatment and prevention of  
CC disorders of bone resorption, e.g. osteoporosis.  
SQ Sequence 15 AA;

Query Match 100.0%; Score 113; DB 20; Length 15;  
Best Local Similarity 100.0%; Pred. No. 6.54e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 etfpkylhydeets 15

QY 1 ETFPKYLHYDEETS 15

## RESULT 2

ID R99951 standard; Protein; 84 AA.  
AC R99951;  
DT 23-APR-1997 (first entry)

DE Mutated OCIF, OCIF-CCR3.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 OS Osteoporosis.  
 FH Key  
 FT Peptide 1..21 Location/Qualifiers  
 FT /note= "Signal peptide"  
 FT Protein 22..84  
 FT /note= "Mature OCIF-CPst"  
 FT WO9626217-A1.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; J00374.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR N-PSDB: T33181.  
 FT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 89; Page 131; 183pp; Japanese.  
 CC This sequence represents a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC sequence represents OCIF-CPst in which amino acids 62-380 of the mature  
 CC OCIF protein are replaced by Leu-Val. These changes are caused by  
 CC the introduction of a restriction site in the DNA encoding this protein.  
 CC The OCIF of the invention has a molecular weight by SDS-PAGE of 60 kD  
 CC under reducing conditions and 120 kD under non-reducing conditions. The  
 CC protein is adsorbed onto cation-exchangers or heparin and its activity is  
 CC lowered after 10 mins at 70 deg.C or 30 mins at 56 deg.C, and is lost  
 CC after 10 mins at 90 deg.C. OCIF is useful in the control of bone  
 CC resorption and therefore in the treatment and prevention of disorders  
 CC of bone resorption, e.g. osteoporosis.  
 SQ Sequence 84 AA;

Query Match 100.0%; Score 113; DB 20; Length 84;  
 Best Local Similarity 100.0%; Pred. No. 6.54e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 22 etfppkylhydeets 36  
 QY 1 ETFPPKYLHYDEETS 15

RESULT 3  
 ID R9947 standard; Protein; 106 AA.  
 AC R9947.  
 DT 23-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-CCR3.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 OS Osteoporosis.  
 FH Key  
 FT Peptide 1..21 Location/Qualifiers  
 FT /note= "Signal peptide"  
 FT Protein 22..106  
 FT /note= "Mature OCIF-CCR3"  
 FT WO9626217-A1.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; J00374.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR N-PSDB: T33177.  
 FT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 77; Page 125; 183pp; Japanese.  
 CC This sequence represents a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This

CC sequence represents OCIF-CCR3 in which amino acids 86-380 of the  
 CC mature OCIF protein are deleted. The OCIF of the invention  
 CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
 CC and 120 kD under non-reducing conditions. The protein is adsorbed onto  
 CC cation-exchangers or heparin and its activity is lowered after 10 mins  
 CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
 CC deg.C. OCIF is useful in the control of bone resorption and therefore  
 CC in the treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 SQ Sequence 106 AA;

Query Match 100.0%; Score 113; DB 20; Length 106;  
 Best Local Similarity 100.0%; Pred. No. 6.54e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 22 etfppkylhydeets 36  
 QY 1 ETFPPKYLHYDEETS 15

RESULT 4  
 ID R9946 standard; Protein; 143 AA.  
 AC R9946.  
 DT 23-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-CCR4.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 OS Osteoporosis.  
 FH Key  
 FT Peptide 1..21 Location/Qualifiers  
 FT /note= "Signal peptide"  
 FT Protein 22..143  
 FT /note= "Mature OCIF-CCR4"  
 FT WO9626217-A1.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; J00374.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR N-PSDB: T33176.  
 FT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 74; Page 124; 183pp; Japanese.  
 CC This sequence represents a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC sequence represents OCIF-CCR4 in which amino acids 123-380 of the  
 CC mature OCIF protein are deleted. The OCIF of the invention  
 CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
 CC and 120 kD under non-reducing conditions. The protein is adsorbed onto  
 CC cation-exchangers or heparin and its activity is lowered after 10 mins  
 CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
 CC deg.C. OCIF is useful in the control of bone resorption and therefore  
 CC in the treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 SQ Sequence 143 AA;

Query Match 100.0%; Score 113; DB 20; Length 143;  
 Best Local Similarity 100.0%; Pred. No. 6.54e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 22 etfppkylhydeets 36  
 QY 1 ETFPPKYLHYDEETS 15

RESULT 5  
 ID R9930 standard; Protein; 145 AA.  
 AC R9930.  
 DT 22-APR-1997 (first entry)  
 DE Osteoclastogenesis inhibitory factor 5.

KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 KW osteoporosis.  
 OS Homo sapiens.  
 FH Key  
 FT Peptide 1..21 Location/Qualifiers  
 FT /note= "Signal peptide"  
 FT Protein 22..145  
 FT /note= "Mature OCIF 5"  
 PN WQ9626217-Al.  
 PD 29-AUG-1996. J00374.  
 PF 20-FEB-1996; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PR (SNOW ) SNOW BRAND MILK PROD CO LTD.  
 PA Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR N-PSDB: T36689.  
 DR DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PT Claim 26; Page 76-77; 183pp; Japanese.  
 CC This sequence represents full length osteoclastogenesis inhibitory  
 factor (OCIF) 5. The OCIF of the invention has a molecular weight by  
 SDS-PAGE of 60 kD under reducing conditions and 120 kD under non-  
 reducing conditions. The protein is adsorbed onto cation-exchangers  
 or heparin and its activity is lowered after 10 mins at 70 deg.C or  
 30 mins at 56 deg.C, and is lost after 10 mins at 90 deg.C. OCIF is  
 useful in the control of bone resorption and therefore in the  
 CC treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 CC Sequence 145 AA;  
 SQ

Query Match 100.0%; Score 113; DB 20; Length 145;  
 Best Local Similarity 100.0%; Pred. No. 6.54e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 22 etfppkylhydeets 36  
 |||||  
 QY 1 ETFPPKYLHYDEETS 15

RESULT 6  
 ID R99929 standard; Protein; 154 AA.  
 AC R99929;  
 DT 22-APR-1997 (first entry)  
 DE Osteoclastogenesis inhibitory factor 4.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 KW osteoporosis.  
 OS Homo sapiens.  
 FH Key  
 FT Peptide 1..21 Location/Qualifiers  
 FT /note= "Signal peptide"  
 FT Protein 22..154  
 FT /note= "Mature OCIF 4"  
 PN WQ9626217-Al.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PR (SNOW ) SNOW BRAND MILK PROD CO LTD.  
 PA Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR N-PSDB: T36688.  
 DR DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PT Claim 23; Page 75; 183pp; Japanese.  
 PS This sequence represents full length osteoclastogenesis inhibitory  
 factor (OCIF) 4. The OCIF of the invention has a molecular weight by  
 SDS-PAGE of 60 kD under reducing conditions and 120 kD under non-  
 reducing conditions. The protein is adsorbed onto cation-exchangers  
 or heparin and its activity is lowered after 10 mins at 70 deg.C or  
 30 mins at 56 deg.C, and is lost after 10 mins at 90 deg.C. OCIF is  
 CC

CC useful in the control of bone resorption and therefore in the  
 CC treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 SQ Sequence 154 AA;

Query Match 100.0%; Score 113; DB 20; Length 154;  
 Best Local Similarity 100.0%; Pred. No. 6.54e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 22 etfppkylhydeets 36  
 |||||  
 QY 1 ETFPPKYLHYDEETS 15

RESULT 7  
 ID R99950 standard; Protein; 187 AA.  
 AC R99950;  
 DT 23-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-CBsp.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 KW osteoporosis.  
 OS Synthetic.  
 FH Key  
 FT Peptide 1..21 Location/Qualifiers  
 FT /note= "Signal peptide"  
 FT Protein 22..187  
 FT /note= "Mature OCIF-CBsp"  
 PN WQ9626217-Al.  
 PD 29-AUG-1996. J00374.  
 PF 20-FEB-1996; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PR (SNOW ) SNOW BRAND MILK PROD CO LTD.  
 PA Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR N-PSDB: T33180.

DR DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PT Claim 86; Page 130-131; 183pp; Japanese.  
 CC This sequence represents a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC sequence represents OCIF-CBsp in which amino acids 167-380 of the mature  
 CC of a restriction site in the DNA encoding this protein. The OCIF of the  
 CC invention has a molecular weight by SDS-PAGE of 60 kD under reducing  
 CC conditions and 120 kD under non-reducing conditions. The protein is  
 CC adsorbed onto cation-exchangers or heparin and its activity is lowered  
 CC after 10 mins at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10  
 CC mins at 90 deg.C. OCIF is useful in the control of bone resorption and  
 CC therefore in the treatment and prevention of disorders of bone  
 CC resorption, e.g. osteoporosis.  
 CC Sequence 187 AA;

Query Match 100.0%; Score 113; DB 20; Length 187;  
 Best Local Similarity 100.0%; Pred. No. 6.54e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 22 etfppkylhydeets 36  
 |||||  
 QY 1 ETFPPKYLHYDEETS 15

RESULT 8  
 ID R99945 standard; Protein; 197 AA.  
 AC R99945;  
 DT 23-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-CDD1.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 KW osteoporosis.  
 OS Synthetic.  
 FH Key  
 FT Peptide 1..21 Location/Qualifiers

FT /note= "Signal peptide"  
 FT Protein 22..197  
 FT /note= "Mature OCIF-CDD1"  
 PN WO9626217-A1.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; J00374.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW ) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR N-PSDB: T33175.  
 PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 71; Page 123; 183pp; Japanese.  
 CC This sequence represents a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC sequence represents OCIF-CDD1 in which amino acids 177-380 of the  
 CC mature OCIF protein are deleted. The OCIF of the invention  
 CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
 CC and 120 kD under non-reducing conditions. The protein is adsorbed onto  
 CC cation-exchangers or heparin and its activity is lowered after 10 mins  
 CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
 CC deg.C. OCIF is useful in the control of bone resorption and therefore  
 CC in the treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 CC Sequence 197 AA;  
 SQ  
 Query Match 100.0%; Score 113; DB 20; Length 197;  
 Best Local Similarity 100.0%; Pred. No. 6.54e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 22 etfppkylhydeets 36  
 QY 1 ETFPPKYLHYDEETS 15  
 RESULT 9  
 ID R99944 standard; Protein; 272 AA.  
 AC R99944;  
 DT 23-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-CDD2.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 KW Osteoporosis.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /note= "Signal peptide"  
 FT Protein 22..272  
 FT /note= "Mature OCIF-CDD2"  
 PN WO9626217-A1.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; J00374.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW ) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR N-PSDB: T33174.  
 PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 68; Page 121-122; 183pp; Japanese.  
 CC This sequence represents a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC sequence represents OCIF-CDD2 in which amino acids 252-380 of the  
 CC mature OCIF protein are deleted. The OCIF of the invention  
 CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
 CC and 120 kD under non-reducing conditions. The protein is adsorbed onto  
 CC cation-exchangers or heparin and its activity is lowered after 10 mins  
 CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
 CC deg.C. OCIF is useful in the control of bone resorption and therefore

CC in the treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 CC Sequence 272 AA;  
 Query Match 100.0%; Score 113; DB 20; Length 272;  
 Best Local Similarity 100.0%; Pred. No. 6.54e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 22 etfppkylhydeets 36  
 QY 1 ETFPPKYLHYDEETS 15  
 RESULT 10  
 ID R99949 standard; Protein; 321 AA.  
 AC R99949;  
 DT 23-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-CSPH.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 KW Osteoporosis.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /note= "Signal peptide"  
 FT Protein 22..321  
 FT /note= "Mature OCIF-CSPH"  
 PN WO9626217-A1.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; J00374.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW ) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR N-PSDB: T33179.  
 PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 83; Page 128-129; 183pp; Japanese.  
 CC This sequence represents a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC OCIF protein are replaced by Ser-Leu-Asp. These changes are caused by  
 CC the introduction of a restriction site in the DNA encoding this protein.  
 CC The OCIF of the invention has a molecular weight by SDS-PAGE of 60 kD  
 CC under reducing conditions and 120 kD under non-reducing conditions. The  
 CC protein is adsorbed onto cation-exchangers or heparin and its activity is  
 CC lowered after 10 mins at 70 deg.C or 30 mins at 56 deg.C, and is lost  
 CC after 10 mins at 90 deg.C. OCIF is useful in the control of bone  
 CC resorption and therefore in the treatment and prevention of disorders  
 CC of bone resorption, e.g. osteoporosis.  
 CC Sequence 321 AA;  
 Query Match 100.0%; Score 113; DB 20; Length 321;  
 Best Local Similarity 100.0%; Pred. No. 6.54e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 22 etfppkylhydeets 36  
 QY 1 ETFPPKYLHYDEETS 15  
 RESULT 11  
 ID R99940 standard; Protein; 326 AA.  
 AC R99940;  
 DT 23-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-DDDI.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 KW Osteoporosis.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /note= "Signal peptide"

FT Protein 22..326  
 FT /note= "Mature OCIF-DDD1"  
 FT Misc\_difference 198..199  
 FT /note= "Position of deletion, delta 178-252"  
 PN WO9626217-A1.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; J00374.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW ) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR N-PSDB: T33170.  
 PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 56; Page 113-114; 183pp; Japanese.  
 CC This sequence represents a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC sequence represents OCIF-DDD1 in which amino acids 178-252 of the  
 CC mature OCIF protein are deleted. The OCIF of the invention  
 CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
 CC and 120 kD under non-reducing conditions. The protein is adsorbed onto  
 CC cation-exchangers or heparin and its activity is lowered after 10 mins  
 CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
 CC deg.C. OCIF is useful in the control of bone resorption and therefore  
 CC in the treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 CC Sequence 326 AA;

Query Match 100.0%; Score 113; DB 20; Length 326;  
 Best Local Similarity 100.0%; Pred. No. 6.54e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 22 etfppkylhydeets 36  
 |||||  
 QY 1 ETFPPKYLHYDEETS 15

RESULT 12  
 ID R99941 standard; Protein: 327 AA.  
 AC R99941;  
 DT 23-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-DDD2.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 OS Osteoporosis.  
 FH Key  
 FT Peptide 1..21  
 FT /note= "Signal peptide"  
 FT Protein 22..327  
 FT /note= "Mature OCIF-DDD2"  
 FT Misc\_difference 273..274  
 FT /note= "Position of deletion, delta 253-326"  
 PN WO9626217-A1.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; J00374.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW ) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR N-PSDB: T33171.

PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 59; Page 115-116; 183pp; Japanese.  
 CC This sequence represents a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC sequence represents OCIF-DDD2 in which amino acids 253-326 of the  
 CC mature OCIF protein are deleted. The OCIF of the invention  
 CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
 CC and 120 kD under non-reducing conditions. The protein is adsorbed onto

CC cation-exchangers or heparin and its activity is lowered after 10 mins  
 CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
 CC deg.C. OCIF is useful in the control of bone resorption and therefore  
 CC in the treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 CC Sequence 327 AA;

Query Match 100.0%; Score 113; DB 20; Length 327;  
 Best Local Similarity 100.0%; Pred. No. 6.54e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 22 etfppkylhydeets 36  
 |||||  
 QY 1 ETFPPKYLHYDEETS 15

RESULT 13  
 ID R99943 standard; Protein: 351 AA.  
 AC R99943;  
 DT 23-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-CC.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 OS Osteoporosis.  
 FH Key  
 FT Peptide 1..21  
 FT /note= "Signal peptide"  
 FT Protein 22..351  
 FT /note= "Mature OCIF-CC"  
 PN WO9626217-A1.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; J00374.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW ) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR N-PSDB: T33173.

PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 65; Page 119-121; 183pp; Japanese.  
 CC This sequence represents a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC sequence represents OCIF-CC in which amino acids 331-380 of the  
 CC mature OCIF protein are deleted. The OCIF of the invention  
 CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
 CC and 120 kD under non-reducing conditions. The protein is adsorbed onto  
 CC cation-exchangers or heparin and its activity is lowered after 10 mins  
 CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
 CC deg.C. OCIF is useful in the control of bone resorption and therefore  
 CC in the treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 CC Sequence 351 AA;

Query Match 100.0%; Score 113; DB 20; Length 351;  
 Best Local Similarity 100.0%; Pred. No. 6.54e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 22 etfppkylhydeets 36  
 |||||  
 QY 1 ETFPPKYLHYDEETS 15

RESULT 14  
 ID R99937 standard; Protein: 359 AA.  
 AC R99937;  
 DT 23-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-DCR2.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 OS Osteoporosis.  
 FH Key  
 FT Peptide 1..21  
 FT /note= "Signal peptide"  
 FT Protein 22..359  
 FT /note= "Mature OCIF-DCR2"  
 FT Misc\_difference 359..359  
 FT /note= "Position of deletion, delta 359-359"  
 PN WO9626217-A1.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; J00374.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW ) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR N-PSDB: T33174.

PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 59; Page 115-116; 183pp; Japanese.  
 CC This sequence represents a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC sequence represents OCIF-DDD2 in which amino acids 253-326 of the  
 CC mature OCIF protein are deleted. The OCIF of the invention  
 CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
 CC and 120 kD under non-reducing conditions. The protein is adsorbed onto

FT Peptide 1..21  
 FT /note= "Signal peptide"  
 FT Protein 22..359  
 FT /note= "Mature OCIF-DCR2"  
 FT Misdifference 63..64  
 FT /note= "Position of deletion, delta 43-84"  
 PN W09526217-A1.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; J00374.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW ) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR N-PSDB: T33167.  
 PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PT Claim 47; Page 107-109; 183pp; Japanese.  
 CC This sequence represents a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC sequence represents OCIF-DCR2 in which amino acids 43-84 of the  
 CC mature OCIF protein are deleted. The OCIF of the invention  
 CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
 CC and 120 kD under non-reducing conditions. The protein is adsorbed onto  
 CC cation-exchangers or heparin and its activity is lowered after 10 mins  
 CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
 CC deg.C. OCIF is useful in the control of bone resorption and therefore  
 CC in the treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 SQ Sequence 359 AA;

Query Match 100.0%; Score 113; DB 20; Length 359;  
 Best Local Similarity 100.0%; Pred. No. 6.54e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 22 etfpkylhydeets 36  
 QY 1 ETFFPKYLHYDEETS 15

RESULT 15  
 ID R99939 standard; Protein: 359 AA.  
 AC R99939;  
 DT 23-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-DCR4.  
 CC Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 CC osteoporosis.  
 CC Synthetic.  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /note= "Signal peptide"  
 FT Protein 22..359  
 FT /note= "Mature OCIF-DCR4"  
 FT Misdifference 143..144  
 FT /note= "Position of deletion, delta 123-164"  
 PN W09526217-A1.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; J00374.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW ) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR N-PSDB: T33169.  
 PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PT Claim 53; Page 111-113; 183pp; Japanese.  
 CC This sequence represents a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC sequence represents OCIF-DCR4 in which amino acids 123-164 of the  
 CC mature OCIF protein are deleted. The OCIF of the invention

CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
 CC and 120 kD under non-reducing conditions. The protein is adsorbed onto  
 CC cation-exchangers or heparin and its activity is lowered after 10 mins  
 CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
 CC deg.C. OCIF is useful in the control of bone resorption and therefore  
 CC in the treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 SQ Sequence 359 AA;

Query Match 100.0%; Score 113; DB 20; Length 359;  
 Best Local Similarity 100.0%; Pred. No. 6.54e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 22 etfpkylhydeets 36  
 QY 1 ETFFPKYLHYDEETS 15

RESULT 16  
 ID R99938 standard; Protein: 360 AA.  
 AC R99938;  
 DT 23-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-DCR3.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 KW osteoporosis.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /note= "Signal peptide"  
 FT Protein 22..360  
 FT /note= "Mature OCIF-DCR3"  
 FT Misdifference 105..106  
 FT /note= "Position of deletion, delta 85-122"  
 PN W09626217-A1.  
 PD 28-AUG-1996.  
 PF 20-FEB-1996; J00374.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW ) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR N-PSDB: T33168.  
 PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PT Claim 50; Page 109-111; 183pp; Japanese.  
 CC This sequence represents a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC sequence represents OCIF-DCR3 in which amino acids 85-122 of the  
 CC mature OCIF protein are deleted. The OCIF of the invention  
 CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
 CC and 120 kD under non-reducing conditions. The protein is adsorbed onto  
 CC cation-exchangers or heparin and its activity is lowered after 10 mins  
 CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
 CC deg.C. OCIF is useful in the control of bone resorption and therefore  
 CC in the treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 SQ Sequence 360 AA;

Query Match 100.0%; Score 113; DB 20; Length 360;  
 Best Local Similarity 100.0%; Pred. No. 6.54e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 22 etfpkylhydeets 36  
 QY 1 ETFFPKYLHYDEETS 15

RESULT 17  
 ID R99924 standard; Protein: 380 AA.  
 AC R99924;  
 DT 22-APR-1997 (first entry)  
 DE Mature osteoclastogenesis inhibitory factor.



KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 KW osteoporosis.  
 OS Homo sapiens.  
 PN WO9626217-A1.  
 PD 29-AUG-1996. J00374.  
 PF 20-FEB-1996; JP-054977.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR N-PSDB; T36685.  
 DR DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 6; Page 62-64; 183pp; Japanese.  
 CC This sequence represents the mature osteoclastogenesis inhibitory  
 CC factor (OCIF) of the invention. The OCIF has a molecular weight by  
 CC SDS-PAGE of 60 kD under reducing conditions and 120 kD under non-  
 CC reducing conditions. The protein is adsorbed onto cation-exchangers  
 CC or heparin and its activity is lowered after 10 mins at 70 deg.C or  
 CC 30 mins at 56 deg.C, and is lost after 10 mins at 90 deg.C. OCIF is  
 CC useful in the control of bone resorption and therefore in the  
 CC treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 SQ Sequence 380 AA;

Query Match 100.0%; Score 113; DB 20; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 6.54e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 etfppkylhydeets 15  
 |||||  
 QY 1 ETFPPKYLHYDEETS 15

RESULT 18  
 ID R99357 standard; Protein: 390 AA.  
 AC R99357;  
 DT 05-MAY-1997 (first entry)  
 DE Human tumour necrosis factor receptor.  
 KW Tumour necrosis factor; TNF; receptor; TNF-beta; ligand; tumour;  
 KW differentiation; immune response; autoimmune disease; inflammation;  
 KW septic shock; graft-versus-host; apoptosis.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /label= sig\_peptide  
 PN WO9628546-A1.  
 PD 19-SEP-1996.  
 PR 15-MAR-1995; U03216.  
 PR 29-MAR-1995; WO-U03216.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Fleischmann RD, Greene JM;  
 DR N-PSDB; T35475.  
 DR New human tumour necrosis factor receptor - used to develop prods.  
 PT for treating e.g. tumours, infection, auto-immune disease, graft  
 PT rejection, cytotoxicity or inflammation  
 PS Claim 14; Fig 1; 59pp; English.  
 CC The receptor binds to TNF, and in particular, TNF-beta.  
 CC The receptor may be used for screening for antagonists and agonists  
 CC of the receptor and for ligands for the receptor. Such agonists may  
 CC be used to inhibit the growth of tumours, to stimulate cellular  
 CC differentiation, to mediate the immune response and anti-viral  
 CC response, to regulate growth and provide resistance to certain  
 CC infections. The antagonists may be used therapeutically, to treat  
 CC autoimmune diseases, inflammation, septic shock, to inhibit graft-  
 CC versus-host reactions, and to prevent apoptosis.  
 SQ Sequence 390 AA;

Query Match 100.0%; Score 113; DB 20; Length 390;

Best Local Similarity 100.0%; Pred. No. 6.54e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 22 etfppkylhydeets 36  
 |||||  
 QY 1 ETFPPKYLHYDEETS 15

RESULT 19  
 ID R99948 standard; Protein: 393 AA.  
 AC R99948;  
 DT 23-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-CBst.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 KW osteoporosis.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /note= "Signal peptide"  
 FT Protein 22..393  
 FT /note= "Mature OCIF-CBst"  
 FT Misc.difference 392  
 FT /label= Gln371leu  
 PN WO9626217-A1.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; J00374.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI; 96-402320/40.  
 DR N-PSDB; T33178.  
 DT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 80; Page 126-128; 183pp; Japanese.  
 CC This sequence represents a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC sequence represents OCIF-CBst in which Gln 371 is substituted by  
 CC Leu and amino acids 373-380 of the mature OCIF protein are deleted.  
 CC These changes are caused by the introduction of a restriction site in  
 CC the DNA encoding this protein. The OCIF of the invention has a  
 CC molecular weight by SDS-PAGE of 60 kD under reducing conditions  
 CC and 120 kD under non-reducing conditions. The protein is adsorbed onto  
 CC cation-exchangers or heparin and its activity is lowered after 10 mins  
 CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
 CC deg.C. OCIF is useful in the control of bone resorption and therefore  
 CC in the treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 SQ Sequence 393 AA;

Query Match 100.0%; Score 113; DB 20; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 6.54e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 22 etfppkylhydeets 36  
 |||||  
 QY 1 ETFPPKYLHYDEETS 15

RESULT 20  
 ID R99942 standard; Protein: 399 AA.  
 AC R99942;  
 DT 23-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-CL.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 KW osteoporosis.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /note= "Signal peptide"  
 FT Protein 22..399  
 FT /note= "Mature OCIF-CL"

PN WO9626217-A1.  
PD 29-AUG-1996.  
PF 20-FEB-1996; J00374.  
PR 20-FEB-1995; JP-054977.  
PR 21-JUL-1995; JP-207508.  
PA (SNOW ) SNOW BRAND MILK PROD CO LTD.  
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
DR WPI: 96-402320/40.  
DR N-PSDB; T33172.  
PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
PT for bone resorption control, esp. treatment of osteoporosis  
PS Claim 62; Page 117-119; 183pp; Japanese.  
CC This sequence represents a mutated version of the full length  
CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
CC sequence represents OCIF-CL in which amino acids 379-380 of the  
CC mature OCIF protein are deleted. The OCIF of the invention  
CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
CC and 120 kD under non-reducing conditions. The protein is adsorbed onto  
CC cation-exchangers or heparin and its activity is lowered after 10 mins  
CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
CC deg.C. OCIF is useful in the control of bone resorption and therefore  
CC in the treatment and prevention of disorders of bone resorption, e.g.  
CC osteoporosis.  
SQ Sequence 399 AA;

Query Match 100.0%; Score 113; DB 20; Length 399;  
Best Local Similarity 100.0%; Pred. No. 6.54e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 22 etfppkylhydeets 36  
QY 1 ETFPPKYLHYDEETS 15

RESULT 21  
ID R99933 standard; Protein; 401 AA.  
AC R99933;  
DE 22-APR-1997 (first entry)  
DE Mutated OCIF, OCIF-C21S.  
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
KW osteoporosis.  
OS Synthetic.  
FH Key Location/Qualifiers  
FT Peptide 1..21  
FT /note= "Signal peptide"  
FT Protein 22..401  
FT /note= "Mature OCIF-C21S"  
FT Misc.difference 277  
FT /label= C21S  
PN WO9626217-A1.  
PD 29-AUG-1996.  
PF 20-FEB-1996; J00374.  
PR 20-FEB-1995; JP-054977.  
PR 21-JUL-1995; JP-207508.  
PA (SNOW ) SNOW BRAND MILK PROD CO LTD.  
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
DR WPI: 96-402320/40.  
DR N-PSDB; T33163.  
PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
PT for bone resorption control, esp. treatment of osteoporosis  
PS Claim 35; Page 98-100; 183pp; Japanese.  
CC This sequence represents a mutated version of the full length  
CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
CC sequence represents OCIF-C21S in which the 21st Cys residue in the  
CC mature OCIF protein is substituted by Ser. The OCIF of the invention  
CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
CC and 120 kD under non-reducing conditions. The protein is adsorbed onto  
CC cation-exchangers or heparin and its activity is lowered after 10 mins  
CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
CC deg.C. OCIF is useful in the control of bone resorption and therefore  
CC in the treatment and prevention of disorders of bone resorption, e.g.

CC osteoporosis.  
SQ Sequence 401 AA;

Query Match 100.0%; Score 113; DB 20; Length 401;  
Best Local Similarity 100.0%; Pred. No. 6.54e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 22 etfppkylhydeets 36  
QY 1 ETFPPKYLHYDEETS 15

RESULT 22  
ID R99932 standard; Protein; 401 AA.  
AC R99932;  
DE 22-APR-1997 (first entry)  
DE Mutated OCIF, OCIF-C20S.  
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
KW osteoporosis.  
OS Synthetic.  
FH Key Location/Qualifiers  
FT Peptide 1..21  
FT /note= "Signal peptide"  
FT Protein 22..401  
FT /note= "Mature OCIF-C20S"  
FT Misc.difference 202  
FT /label= C20S  
PN WO9626217-A1.  
PD 29-AUG-1996.  
PF 20-FEB-1996; J00374.  
PR 20-FEB-1995; JP-054977.  
PR 21-JUL-1995; JP-207508.  
PA (SNOW ) SNOW BRAND MILK PROD CO LTD.  
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
DR WPI: 96-402320/40.  
DR N-PSDB; T33162.

PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
PT for bone resorption control, esp. treatment of osteoporosis  
PS Claim 32; Page 96-98; 183pp; Japanese.  
CC This sequence represents a mutated version of the full length  
CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
CC sequence represents OCIF-C20S in which the 20th Cys residue in the  
CC mature OCIF protein is substituted by Ser. The OCIF of the invention  
CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions  
CC and 120 kD under non-reducing conditions. The protein is adsorbed onto  
CC cation-exchangers or heparin and its activity is lowered after 10 mins  
CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
CC deg.C. OCIF is useful in the control of bone resorption and therefore  
CC in the treatment and prevention of disorders of bone resorption, e.g.  
CC osteoporosis.  
SQ Sequence 401 AA;

Query Match 100.0%; Score 113; DB 20; Length 401;  
Best Local Similarity 100.0%; Pred. No. 6.54e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 22 etfppkylhydeets 36  
QY 1 ETFPPKYLHYDEETS 15

RESULT 23  
ID R99935 standard; Protein; 401 AA.  
AC R99935;  
DE 22-APR-1997 (first entry)  
DE Mutated OCIF, OCIF-C23S.  
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
KW osteoporosis.  
OS Synthetic.  
FH Key Location/Qualifiers  
FT Peptide 1..21  
FT /note= "Signal peptide"

```

FT Protein 22..401
FT /note= "Mature OCIF-C23S"
FT Misc_difference 400
FT /label= C23S
PN W09626217-A1.
PD 29-AUG-1996.
PF 20-FEB-1996; J00374.
PR 20-FEB-1995; JP-054977.
PR 21-JUL-1995; JP-207508.
PA (SNOW ) SNOW BRAND MILK PROD CO LTD.
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;
DR WPI: 96-402320/40.
DR N-PSDB: T33165.
PT DNA encoding osteoclastogenesis inhibitory factor protein - useful
PT for bone resorption control, esp. treatment of osteoporosis
PS Claim 41: Page 103-105; 183pp; Japanese.
CC This sequence represents a mutated version of the full length
CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This
CC sequence represents OCIF-C23S in which the 23rd Cys residue in the
CC mature OCIF protein is substituted by Ser. The OCIF of the invention
CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions
CC and 120 kD under non-reducing conditions. The protein is adsorbed onto
CC cation-exchangers or heparin and its activity is lowered after 10 mins
CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90
CC deg.C. OCIF is useful in the control of bone resorption and therefore
CC in the treatment and prevention of disorders of bone resorption, e.g.
CC osteoporosis.
SQ Sequence 401 AA;

Query Match 100.0%; Score 113; DB 20; Length 401;
Best Local Similarity 100.0%; Pred. No. 6.54e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 22 etfppkylhydeets 36
Qy 1 ETFPKYLHYDEETS 15

RESULT 25
ID R99931 standard; Protein; 401 AA.
AC R99931;
DT 22-APR-1997 (first entry)
DE Mutated OCIF, OCIF-C19S.
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;
KW osteoporosis.
OS Synthetic.
FH Key Location/Qualifiers
FT Peptide 1..21
FT /note= "signal peptide"
FT Protein 22..401
FT /note= "Mature OCIF-C19S"
FT Misc_difference 195
FT /label= C19S
PN W09626217-A1.
PD 29-AUG-1996.
PE 20-FEB-1996; J00374.
PR 20-FEB-1995; JP-054977.
PR 21-JUL-1995; JP-207508.
PA (SNOW ) SNOW BRAND MILK PROD CO LTD.
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;
DR WPI: 96-402320/40.
DR N-PSDB: T33161.
PT DNA encoding osteoclastogenesis inhibitory factor protein - useful
PT for bone resorption control, esp. treatment of osteoporosis
PS Claim 29: Page 94-96; 183pp; Japanese.
CC This sequence represents a mutated version of the full length
CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This
CC sequence represents OCIF-C19S in which the 19th Cys residue in the
CC mature OCIF protein is substituted by Ser. The OCIF of the invention
CC has a molecular weight by SDS-PAGE of 60 kD under reducing conditions
CC and 120 kD under non-reducing conditions. The protein is adsorbed onto
CC cation-exchangers or heparin and its activity is lowered after 10 mins
CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90
CC deg.C. OCIF is useful in the control of bone resorption and therefore
CC in the treatment and prevention of disorders of bone resorption, e.g.
CC osteoporosis.
SQ Sequence 401 AA;

Query Match 100.0%; Score 113; DB 20; Length 401;
Best Local Similarity 100.0%; Pred. No. 6.54e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 22 etfppkylhydeets 36
Qy 1 ETFPKYLHYDEETS 15

RESULT 26
ID R99925 standard; Protein; 401 AA.
AC R99925;
DT 22-APR-1997 (first entry)
DE Full length osteoclastogenesis inhibitory factor.
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;
KW osteoporosis.

```

OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /note= "Signal peptide"  
 ET Protein 22..401  
 FT /note= "Mature OCIF, Claim 6"  
 PN WO9626217-A1.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; J00374.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR N-PSDB: T36685.  
 PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PT Disclosure: Page 64-66; 183pp; Japanese.  
 CC This sequence represents the full length osteoclastogenesis inhibitory  
 CC factor (OCIF) of the invention. The OCIF has a molecular weight by  
 CC SDS-PAGE of 60 kD under reducing conditions and 120 kD under non-  
 CC reducing conditions. The protein is adsorbed onto cation-exchangers  
 CC or heparin and its activity is lowered after 10 mins at 70 deg.C or  
 CC 30 mins at 56 deg.C and is lost after 10 mins at 90 deg.C. OCIF is  
 CC useful in the control of bone resorption and therefore in the  
 CC treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 SQ Sequence 401 AA;

Query Match 100.0%; Score 113; DB 20; Length 401;  
 Best Local Similarity 100.0%; Pred.No. 6.54e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 22 etfppkylhydeets 36  
 QY 1 ETfPPKYLHYDEETS 15  
 |||||  
 |||||

Search completed: Wed Dec 3 10:07:20 1997  
 Job time : 6 secs.

\*\*\*\*\*

(TM)

\*\*\*\*\*

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by Intelligenetics, Inc.

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Tue Dec 2 14:52:31 1997; MasPar time 2.86 Seconds  
151.367 Million cell updates/sec  
Circular output not generated.

Title: >US-08-915-004-7  
Description: (1-15) from US08915004.pap  
Perfect Score: 113  
Sequence: 1 ETPPPKYLHYDEETS 15

Scoring table: PAM 150  
Gap 15

Searched: 91006 seqs, 28888923 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pi-51  
1:ann1 2:ann2 3:ann3 4:ann4 5:unann1 6:unann2 7:unann3  
8:unann4 9:unann5 10:unann6 11:unann7 12:unann8  
13:unann9 14:unann10 15:unenc 16:unrev

Statistics: Mean 26.254; Variance 43.975; scale 0.597

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	ID	Description	Pred. No.
1	67	59.3	1932	12	probable membrane pr	3.59e-01
2	65	57.5	386	9	histidine decarboxyl	7.73e-01
3	62	54.9	592	16	beta-fructofuranosid	2.38e-00
4	62	54.9	592	5	beta-fructofuranosid	2.38e-00
5	60	53.1	37	12	MHC class I protein	4.95e-00
6	60	53.1	525	5	cytochrome-c oxidase	4.95e-00
7	58	51.3	294	11	hypothetical protein	1.01e-01
8	58	51.3	643	10	72k crystal protein	1.01e-01
9	56	49.6	203	9	hypothetical protein	2.05e-01
10	56	49.6	273	10	hypothetical protein	2.05e-01
11	56	49.6	330	9	hypothetical protein	2.05e-01
12	56	49.6	525	5	cytochrome-c oxidase	2.05e-01
13	56	49.6	678	9	hypothetical protein	2.05e-01
14	56	49.6	1056	12	probable membrane pr	2.05e-01
15	55	48.7	353	1	farnesyl-pyrophosphat	2.89e-01
16	55	48.7	463	14	ribulose-bisphosphat	2.89e-01
17	55	48.7	565	16	malate oxidoreductase	2.89e-01
18	55	48.7	596	1	acetylcholinesterase	2.89e-01
19	55	48.7	599	1	acetylcholinesterase	2.89e-01
20	55	48.7	604	7	probable amino acid	2.89e-01
21	55	48.7	759	12	arylphorin precursor	2.89e-01

22	54	47.8	154	3	EYNKHR	egg-lysin precursor	4.06e+01
23	54	47.8 <td>155</td> <td>7</td> <td>B35960 <td>egg-lysin precursor <th>4.06e+01</th> </td></td>	155	7	B35960 <td>egg-lysin precursor <th>4.06e+01</th> </td>	egg-lysin precursor <th>4.06e+01</th>	4.06e+01
24	54	47.8 <td>169</td> <td>12</td> <td>S40994 <td>hypothetical protein <th>4.06e+01</th> </td></td>	169	12	S40994 <td>hypothetical protein <th>4.06e+01</th> </td>	hypothetical protein <th>4.06e+01</th>	4.06e+01
25	54	47.8 <td>340</td> <td>10</td> <td>E64435 <td>hypothetical protein <th>4.06e+01</th> </td></td>	340	10	E64435 <td>hypothetical protein <th>4.06e+01</th> </td>	hypothetical protein <th>4.06e+01</th>	4.06e+01
26	54	47.8 <td>466</td> <td>6</td> <td>S07295 <td>ribulose-bisphosphat <th>4.06e+01</th> </td></td>	466	6	S07295 <td>ribulose-bisphosphat <th>4.06e+01</th> </td>	ribulose-bisphosphat <th>4.06e+01</th>	4.06e+01
27	54	47.8 <td>652</td> <td>9</td> <td>S25265 <td>outer membrane prote <th>4.06e+01</th> </td></td>	652	9	S25265 <td>outer membrane prote <th>4.06e+01</th> </td>	outer membrane prote <th>4.06e+01</th>	4.06e+01
28	54	47.8 <td>1036</td> <td>12</td> <td>S55984 <td>probable membrane pr <th>4.06e+01</th> </td></td>	1036	12	S55984 <td>probable membrane pr <th>4.06e+01</th> </td>	probable membrane pr <th>4.06e+01</th>	4.06e+01
29	53	46.9 <td>20</td> <td>14</td> <td>C49164 <td>chromogranin-B - rat <th>5.68e+01</th> </td></td>	20	14	C49164 <td>chromogranin-B - rat <th>5.68e+01</th> </td>	chromogranin-B - rat <th>5.68e+01</th>	5.68e+01
30	53	46.9 <td>190</td> <td>8</td> <td>I42522 <td>A52R protein - vacci <th>5.68e+01</th> </td></td>	190	8	I42522 <td>A52R protein - vacci <th>5.68e+01</th> </td>	A52R protein - vacci <th>5.68e+01</th>	5.68e+01
31	53	46.9 <td>190</td> <td>8</td> <td>J01790 <td>SalF15R protein - va <th>5.68e+01</th> </td></td>	190	8	J01790 <td>SalF15R protein - va <th>5.68e+01</th> </td>	SalF15R protein - va <th>5.68e+01</th>	5.68e+01
32	53	46.9 <td>400</td> <td>9</td> <td>S32879 <td>lipA protein - Neiss <th>5.68e+01</th> </td></td>	400	9	S32879 <td>lipA protein - Neiss <th>5.68e+01</th> </td>	lipA protein - Neiss <th>5.68e+01</th>	5.68e+01
33	53	46.9 <td>489</td> <td>10</td> <td>J00767 <td>beta-galactosidase ( <th>5.68e+01</th> </td></td>	489	10	J00767 <td>beta-galactosidase ( <th>5.68e+01</th> </td>	beta-galactosidase ( <th>5.68e+01</th>	5.68e+01
34	53	46.9 <td>491</td> <td>10</td> <td>S06762 <td>beta-galactosidase ( <th>5.68e+01</th> </td></td>	491	10	S06762 <td>beta-galactosidase ( <th>5.68e+01</th> </td>	beta-galactosidase ( <th>5.68e+01</th>	5.68e+01
35	53	46.9 <td>521</td> <td>12</td> <td>S55317 <td>cytochrome P450 1A - <th>5.68e+01</th> </td></td>	521	12	S55317 <td>cytochrome P450 1A - <th>5.68e+01</th> </td>	cytochrome P450 1A - <th>5.68e+01</th>	5.68e+01
36	53	46.9	609	7	S45930 <td>probable amino acid <th>5.68e+01</th> </td>	probable amino acid <th>5.68e+01</th>	5.68e+01
37	53	46.9	3562	13	A47171 <td>chondroitin sulfate <th>5.68e+01</th> </td>	chondroitin sulfate <th>5.68e+01</th>	5.68e+01
38	52	46.0	146	16	JC4902 <td>ionized calcium bind <th>7.92e-01</th> </td>	ionized calcium bind <th>7.92e-01</th>	7.92e-01
39	52	46.0	147	14	I55617 <td>allograft inflammato <th>7.92e-01</th> </td>	allograft inflammato <th>7.92e-01</th>	7.92e-01
40	52	46.0	150	14	I46004 <td>calpastatin - bovine <th>7.92e-01</th> </td>	calpastatin - bovine <th>7.92e-01</th>	7.92e-01
41	52	46.0	159	12	S28818 <td>transforming protein <th>7.92e-01</th> </td>	transforming protein <th>7.92e-01</th>	7.92e-01
42	52	46.0	200	7	A37145 <td>ribosomal protein S4 <th>7.92e-01</th> </td>	ribosomal protein S4 <th>7.92e-01</th>	7.92e-01
43	52	46.0	235	14	I56987 <td>BART-1 - rat <th>7.92e-01</th> </td>	BART-1 - rat <th>7.92e-01</th>	7.92e-01
44	52	46.0	798	16	S62405 <td>hypothetical protein <th>7.92e-01</th> </td>	hypothetical protein <th>7.92e-01</th>	7.92e-01
45	52	46.0	1252	6	S21178 <td>botulinum neurotoxin <th>7.92e-01</th> </td>	botulinum neurotoxin <th>7.92e-01</th>	7.92e-01

## ALIGNMENTS

RESULT 1  
ENTRY #type complete  
TITLE probable membrane protein YLR422w - yeast (Saccharomyces cerevisiae)  
ALTERNATE\_NAMES hypothetical protein L9576.7  
ORGANISM #formal\_name Saccharomyces cerevisiae  
DATE 05-May-1995 #sequence\_revision 01-Sep-1995 #text\_change 19-Jul-1996

ACCESSIONS S53409  
REFERENCE S53409  
#authors Favell, A.  
#submission submitted to the EMBL Data Library, February 1995  
#description The sequence of S. cerevisiae cosmid 9576.  
#accession S53409  
##molecule\_type DNA  
##residues 1-1932 ##label FAV  
##cross-references EMBL:U20939  
##experimental\_source strain S288C (AB972)

GENETICS

#map\_position 12R  
KEYWORDS transmembrane protein

FEATURE  
841-857 #domain transmembrane #status predicted #label TM1  
1079-1095 #domain transmembrane #status predicted #label TM2  
1157-1173 #domain transmembrane #status predicted #label TM3  
SUMMARY #length 1932 #molecular-weight 221561 #checksum 5013

Query Match 59.3%; Score 67; DB 12; Length 1932;  
Best Local Similarity 46.2%; Pred. No. 3.59e-01;  
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 95 fprkywhfeaea 107

Qy 3 FPPKYLHYDEETS 15

RESULT 2

ENTRY #type complete  
TITLE histidine decarboxylase (EC 4.1.1.22) - Vibrio anguillarum  
ORGANISM #formal\_name Vibrio anguillarum  
DATE 16-Feb-1995 #sequence\_revision 12-May-1995 #text\_change 22-Nov-1996  
ACCESSIONS S60898; S49218  
REFERENCE S60898  
#authors Tolmashy, M.E.; Actis, L.A.; Crosa, J.H.

```

#journal Mol. Microbiol. (1995) 15:87-95
#title A histidine decarboxylase gene encoded by the Vibrio
#residues anguillarum plasmid pJMI is essential for virulence:
#accession histamine is a precursor in the biosynthesis of
#status S60898
#molecule_type preliminary
#residues 1-386 ##label T02
#cross-references EMBL:Z33880
KEYWORDS carbon-carbon lyase; carboxy-lyase; pyridoxal phosphate
FEATURE 233
#binding_site pyridoxal phosphate (Lys) (covalent)
#status predicted
#length 386 #molecular-weight 44260 #checksum 4338
Query Match 57.5%; Score 65; DB 9; Length 386;
Best Local Similarity 57.1%; Pred. No. 7.73e-01;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Db 106 eiftpkwyseet 119
| | | | |
QY 1 ETEPPKYLHYDEE 14

RESULT 3
ENTRY S61503 #type complete
TITLE beta-fructofuranosidase (EC 3.2.1.26) - carrot
ORGANISM #formal_name Daucus carota #common_name carrot
DATE 27-Oct-1996 #sequence_revision 27-Oct-1996 #text_change
27-Oct-1996
ACCESSIONS S61503
REFERENCE S61503
#authors Ramloeh-Lorenz, K.; Knudsen, S.; Sturm, A.
#journal Plant J. (1993) 4:545-554
#title Molecular characterization of the gene for carrot cell wall
#accession beta-fructosidase.
#status S61503
#residues preliminary
#cross-references EMBL:X69321
SUMMARY #length 592 #molecular-weight 66871 #checksum 4674

Query Match 54.9%; Score 62; DB 16; Length 592;
Best Local Similarity 61.5%; Pred. No. 2.38e+00;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

435 efpdkwleydae 447
| | | | |
QY 1 ETEPPKYLHYDEE 13

RESULT 4
ENTRY JQ0991 #type complete
TITLE beta-fructofuranosidase (EC 3.2.1.26) precursor - carrot
ALTERNATE_NAMES beta-fructosidase; vacuolar invertase
ORGANISM #formal_name Daucus carota #common_name carrot
DATE 10-Mar-1994 #sequence_revision 10-Mar-1994 #text_change
10-Mar-1994
ACCESSIONS JQ0991
REFERENCE JQ0991
#authors Sturm, A.; Chrispeels, M.J.
#journal Plant Cell (1990) 2:1107-1119
#title cDNA cloning of carrot extracellular beta-fructosidase and
its expression in response to wounding and bacterial
infection.
#cross-references MUID:93005650
#accession JQ0991
#status preliminary
#molecule_type mRNA
#residues 1-592 ##label STU
#cross-references GB:M58362
CLASSIFICATION #superfamily beta-fructofuranosidase

```

```

KEYWORDS glycosidase; hydrolase
SUMMARY #length 592 #molecular-weight 66813 #checksum 3928

Query Match 54.9%; Score 62; DB 5; Length 592;
Best Local Similarity 61.5%; Pred. No. 2.38e+00;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 435 efpdkwleydae 447
| | | | |
QY 1 ETEPPKYLHYDEE 13

RESULT 5
ENTRY I50036 #type fragment
TITLE MHC class I protein - axolotl (fragment)
ORGANISM #formal_name Ambystoma mexicanum #common_name axolotl
DATE 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change
13-Sep-1996
ACCESSIONS I50036
REFERENCE I50012
#authors Grossberger, D.; Parham, P.
#journal Immunogenetics (1992) 36:166-174
#title Reptilian class I major histocompatibility complex genes
reveal conserved elements in class I structure.
#cross-references MUID:92307752
#accession I50036
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-37 ##label GRO
#cross-references GB:M81089; NID:g210935; CDS_PID:g555414
GENETICS
#note gene name A2
SUMMARY #length 37 #checksum 4645

Query Match 53.1%; Score 60; DB 12; Length 37;
Best Local Similarity 77.8%; Pred. No. 4.95e-00;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 3 kylgygeet 11
| | | | |
QY 6 KYLHYDEET 14

RESULT 6
ENTRY S26034 #type complete
TITLE cytochrome-c oxidase (EC 1.9.3.1) chain I - Caenorhabditis
ORGANISM elegans mitochondrion (SGC4)
DATE #formal_name mitochondrion Caenorhabditis elegans
12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change
17-Nov-1995
ACCESSIONS S26034; S25807
REFERENCE S26014
#authors Okimoto, R.; Macfarlane, J.L.; Clary, D.O.; Wolstenholme,
D.R.
#journal Genetics (1992) 130:471-498
#title The mitochondrial genomes of two nematodes, Caenorhabditis
elegans and Ascaris suum.
#cross-references MUID:92201635
#accession S26034
#molecule_type DNA
#residues 1-525 ##label OKI
#cross-references EMBL:X54252
#note the authors translated the initiation codon ATT for
residue 1 as Ile
REFERENCE S13139
#authors Okimoto, R.; Macfarlane, J.L.; Wolstenholme, D.R.
#journal Nucleic Acids Res (1990) 18:6113-6118
#title Evidence for the frequent use of ATG as the translation
initiation codon of mitochondrial protein genes in the
nematodes, Ascaris suum and Caenorhabditis elegans.
#accession S25807
#molecule_type DNA
#residues 1-25 ##label OK2

```

```

##cross-references EMBL:X54252
##note the authors translated the initiation codon ATT for
        residue 1 as Ile

GENETICS
#gene      COI
#genome    mitochondrion
#genetic_code  SGC4
#start_codon ATT
CLASSIFICATION #superfamily cytochrome-c oxidase chain I; cytochrome-c
                oxidase chain I homology
                copper binding; electron transfer; heme; membrane-associated
                complex; mitochondrion; oxidative phosphorylation;
                oxidoreductase; respiratory chain; transmembrane protein

KEYWORDS

FEATURE
        19-464
SUMMARY
        #domain cytochrome-c oxidase chain I homology #label COI
        #length 525 #molecular-weight 58463 #checksum 9978

        Query Match      53.1%; Score 60; DB 5; Length 525;
        Best Local Similarity 60.0%; Pred. No. 4.95e+00;
        Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

443 fprkyidydpd 452
      |||||
      3 FPPKYLHYDE 12

RESULT 7
ENTRY S67612 #type complete
TITLE hypothetical protein YDL076c - yeast (Saccharomyces
        cerevisiae)
ALTERNATE_NAMES hypothetical protein D2476
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change
        12-Jul-1996
ACCESSIONS S67612
REFERENCE S67608
#authors Wambutt, R.; Wedler, H.; Wedler, E.; Scharfe, M.
#submission submitted to the Protein Sequence Database, July 1996
#accession S67612
#molecule_type DNA
#residues 1-294 #label WAM
##cross-references EMBL:Z74124
##experimental_source strain S288C

GENETICS
#map_position 4L
#map_length 294 #molecular-weight 33811 #checksum 9485
SUMMARY

        Query Match      51.3%; Score 58; DB 11; Length 294;
        Best Local Similarity 45.5%; Pred. No. 1.01e+01;
        Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

46 sfpeqymhyqn 56
      :|||:|:|:|
      2 TFPKYLHYDE 12

RESULT 8
ENTRY A43647 #type complete
TITLE 72K crystal protein - Bacillus thuringiensis subsp.
        israelensis
ORGANISM #formal_name Bacillus thuringiensis subsp. israelensis
DATE 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change
        01-Mar-1996
ACCESSIONS A43647; A32256
REFERENCE A43647
#authors Donovan, W.P.; Dankocsik, C.; Gilbert, M.P.
#journal J. Bacteriol. (1988) 170:4732-4738
#title Molecular characterization of a gene encoding a 72-kilodalton
        mosquito-toxic crystal protein from Bacillus thuringiensis
        subsp. israelensis.

#accession A43647 preliminary
##status preliminary
##molecule_type DNA

```

```

##residues 1-643 #label DON
##cross-references GB:M31737
REFERENCE A32256
#authors Adams, L.F.; Visick, J.E.; Whiteley, H.R.
#journal J. Bacteriol. (1989) 171:521-530
#title A 20-kilodalton protein is required for efficient production
        of the Bacillus thuringiensis subsp. israelensis
        27-kilodalton crystal protein in Escherichia coli.

#cross-references MUID:89123065
#accession A32256
#status preliminary; not compared with conceptual translation
#molecule_type DNA
#residues 566-643 #label ADA

GENETICS
#gene crypD
SUMMARY #length 643 #molecular-weight 72348 #checksum 6237

        Query Match      51.3%; Score 58; DB 10; Length 643;
        Best Local Similarity 70.0%; Pred. No. 1.01e+01;
        Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 413 tipnpylayd 422
      |||||
      2 TFPKYLHYD 11

RESULT 9
ENTRY S01180 #type complete
TITLE hypothetical protein N-1 - Chlamydia trachomatis plasmid
        pUGV440
ORGANISM #formal_name Chlamydia trachomatis
DATE 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change
        18-Jun-1993
ACCESSIONS S01180
REFERENCE S00788
#authors Hatt, C.; Ward, M.E.; Clarke, I.N.
#journal Nucleic Acids Res. (1988) 16:4053-4067
#title Analysis of the entire nucleotide sequence of the cryptic
        plasmid of Chlamydia trachomatis serovar L1. Evidence for
        involvement in DNA replication.

#cross-references MUID:88233998
#accession S01180
#molecule_type DNA
#residues 1-203 #label HAT
##cross-references EMBL:X06707

GENETICS
#genome plasmid
SUMMARY #length 203 #molecular-weight 23076 #checksum 8385

        Query Match      49.6%; Score 56; DB 9; Length 203;
        Best Local Similarity 41.7%; Pred. No. 2.05e+01;
        Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 100 typqkfmhflge 111
      |||||:|:|:|
      2 TFPKYLHYDE 13

RESULT 10
ENTRY S07880 #type complete
TITLE hypothetical protein 1 - Methanobacterium thermoautotrophicum
        plasmid pME2001
ORGANISM #formal_name Methanobacterium thermoautotrophicum
DATE 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change
        18-Jun-1993
ACCESSIONS S07880
REFERENCE S07880
#authors Bokranz, M.; Klein, A.; Meile, L.
#journal Nucleic Acids Res. (1990) 18:363
#title Complete nucleotide sequence of plasmid pME2001 of
        Methanobacterium thermoautotrophicum (Marburg).

#cross-references MUID:90221821
#accession S07880

```

```

##molecule_type DNA
##residues 1-273 ##label BOX
##cross-references EMBL:X17205
GENETICS
#genome
#summary
#length 273 #molecular-weight 30744 #checksum 8715
Query Match 49.6%; Score 56; DB 10; Length 273;
Best Local Similarity 58.3%; Pred. No. 2.05e+01;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
Db 223 spkylfysatg 234
:||||| |
QY 4 PPKYLHYDETS 15

RESULT 11
ENTRY
TITLE
H37386 #type complete
hypothetical protein 8 - Chlamydia trachomatis (serotype D)
plasmids
#formal_name Chlamydia trachomatis
#formal_name Chlamydia trachomatis
17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change
01-Dec-1995
ACCESSIONS H37386; S01920; S31170
REFERENCE A37386
#authors Comanducci, M.; Ricci, S.; Cevenini, R.; Ratti, G.
#journal Plasmid (1990) 23:149-154
#title Diversity of the Chlamydia trachomatis common plasmid in
biovars with different pathogenicity.
#accession H37386
##status preliminary
##molecule_type DNA
##residues 1-330 ##label COM1
##cross-references GB:J03321
REFERENCE S01920
#authors Comanducci, M.; Ricci, S.; Ratti, G.
#journal Mol. Microbiol. (1988) 2:531-538.
#title The structure of a plasmid of Chlamydia trachomatis believed
to be required for growth within mammalian cells.
#cross-references MUID:89013895
#accession S01920
##molecule_type DNA
##residues 1-330 ##label COM2
##cross-references EMBL:X07547
##experimental_source plasmid pCHL1
REFERENCE S31169
#authors Ricci, S.; Cevenini, R.; Cosco, E.; Comanducci, M.; Ratti,
G.; Scariato, V.
#journal Mol. Gen. Genet. (1993) 237:318-326
#title Transcriptional analysis of the Chlamydia trachomatis plasmid
pCT identifies temporally regulated transcripts, anti-sense
RNA and sigma(70)-selected promoters.
#accession S31170
##status not compared with conceptual translation
##molecule_type DNA
##residues 1-330 ##label RIC
##experimental_source plasmid pCT
GENETICS
#genome
#summary
#length 330 #molecular-weight 37791 #checksum 1740
Query Match 49.6%; Score 56; DB 9; Length 330;
Best Local Similarity 41.7%; Pred. No. 2.05e+01;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
Db 225.typpqkfmhflqe 236
:|:|:|:|:|
QY 2 TFPKKYLHYDEE 13

RESULT 12
ENTRY
TITLE
S26022 #type complete
cytochrome-c oxidase (EC 1.9.3.1) chain I - pig roundworm

```

```

ORGANISM mitochondrion (SGC4)
#formal_name mitochondrion Ascaris suum #common_name pig
roundworm
DATE 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change
17-Nov-1995
ACCESSIONS S26022; S25795
REFERENCE S26014
#authors Okimoto, R.; Macfarlane, J.L.; Clary, D.O.; Wolstenholme,
D.R.
#journal Genetics (1992) 130:471-498
#title The mitochondrial genomes of two nematodes, Caenorhabditis
elegans and Ascaris suum.
#cross-references MUID:92201635
#accession S26022
##molecule_type DNA
##residues 1-525 ##label OKI
##cross-references EMBL:X54253
##note the authors translated the initiation codon ATT for
residue 1 as Ile
REFERENCE S13139
#authors Okimoto, R.; Macfarlane, J.L.; Wolstenholme, D.R.
#journal Nucleic Acids Res. (1990) 18:6113-6118
#title Evidence for the frequent use of TTG as the translation
initiation codon of mitochondrial protein genes in the
nematodes, Ascaris suum and Caenorhabditis elegans.
#accession S25795
##molecule_type DNA
##residues 1-25 ##label OK2
##cross-references EMBL:X54253
##note the authors translated the initiation codon ATT for
residue 1 as Ile
GENETICS
#gene COI
#genome mitochondrion
#genetic_code SGC4
#start_codon ATT
CLASSIFICATION #superfamily cytochrome-c oxidase chain I; cytochrome-c
oxidase chain I homology
KEYWORDS copper binding; electron transfer; heme; membrane-associated
complex; mitochondrion; oxidative phosphorylation;
oxidoreductase; respiratory chain; transmembrane protein
FEATURE
19-464 #domain cytochrome-c oxidase chain I homology #label COI
SUMMARY #length 525 #molecular-weight 58529 #checksum 1290
Query Match 49.6%; Score 56; DB 5; Length 525;
Best Local Similarity 50.0%; Pred. No. 2.05e+01;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
Db 443 yprkyldypd 452
:| | | |
QY 3 FPPKYLHYDE 12

RESULT 13
ENTRY
TITLE
S40822 #type complete
hypothetical protein f678 - Escherichia coli
#formal_name Escherichia coli
#formal_name Escherichia coli
19-May-1994 #sequence_revision 01-Sep-1995 #text_change
01-Sep-1995
ACCESSIONS S40822
REFERENCE S40802
#authors Plunkett III, G.; Burland, V.; Daniels, D.L.; Blattner, F.R.
#journal Nucleic Acids Res. (1993) 21:3391-3398
#title Analysis of the Escherichia coli genome. III. DNA sequence of
the region from 87.2 to 89.2 minutes.
#accession S40822
##status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-678 ##label PLU
##cross-references EMBL:L19201
##note the nucleotide sequence was submitted to the EMBL Data

```



```

SUMMARY      Library, October 1993
#length 678 #molecular-weight 77248 #checksum 649
Query Match  49.6%; Score 56; DB 9; Length 678;
Best Local Similarity 40.0%; Pred. No. 2.05e+01;
Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 589 pflhyedda 598
|:||||:
QY 5 PPKYLHYDEET 14

RESULT 14
ENTRY S55151 #type complete
TITLE probable membrane protein YN1132w - yeast (Saccharomyces cerevisiae)

ALTERNATE_NAMES
hypothetical protein JTA1056; hypothetical protein N1216;
hypothetical protein N1858

ORGANISM #formal_name Saccharomyces cerevisiae
DATE 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change
#accession S55151; S59256; S63077
#residues S55136
#authors Mallet, L.; Bussereau, F.; Jacquet, M.
#submission submitted to the EMBL Data Library, November 1994
#description A 43.5 kb fragment of the chromosome XIV.
#accession S55151
#molecule_type DNA
#residues 1-1056 #label MAL
#cross-references EMBL:Z46843

REFERENCE S59241
#authors Mallet, L.; Bussereau, F.; Jacquet, M.
#journal Yeast (1995) 11:1195-1209
#title A 43.5 kb segment of yeast chromosome XIV, which contains MFA2, MFP2, CAP/SRV2, NAM9, FKB1/PPRI/RBP1, MOM22 and CP11, predicts an adenosine deaminase gene and 14 new open reading frames.
#accession S59256
#status nucleic acid sequence not shown; translation not shown
#residues 1-1056 #label MAW
#cross-references EMBL:Z46843
#note the nucleotide sequence was submitted to the EMBL Data Library, November 1994

REFERENCE S63069
#authors Mallet, L.; Bussereau, F.; Jacquet, M.
#submission submitted to the Protein Sequence Database, April 1996
#accession S63077
#molecule_type DNA
#residues 1-1056 #label MAF
#cross-references EMBL:Z71408
#experimental_source strain S288C

GENETICS
#map_position 14L
#keywords transmembrane protein
FEATURE
381-397 #domain transmembrane #status predicted #label TMM
SUMMARY #length 1056 #molecular-weight 119346 #checksum 3315

Query Match 49.6%; Score 56; DB 12; Length 1056;
Best Local Similarity 85.7%; Pred. No. 2.05e+01;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 716 pphylhy 722
|:|||||
QY 4 PPKYLHY 10

RESULT 15
ENTRY A35726 #type complete
TITLE farnesyl-pyrophosphate synthetase - human
ALTERNATE_NAMES
prenyltransferase and farnesyl-diphosphate synthase
CONTAINS dimethylallyltransferase (EC 2.5.1.1);

```

```

ORGANISM geranyltransferase (EC 2.5.1.10)
DATE #formal_name Homo sapiens #common_name man
30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change
06-Sep-1996
ACCESSIONS A35726; A33415
REFERENCE #authors Wilkin, D.J.; Kutsunai, S.Y.; Edwards, P.A.
#journal J. Biol. Chem. (1990) 265:4607-4614
#title Isolation and sequence of the human farnesyl pyrophosphate
synthetase cDNA. Coordinate regulation of the mRNAs for
farnesyl pyrophosphate synthetase,
3-hydroxy-3-methylglutaryl coenzyme A reductase, and
ester.
#cross-references MIM:90170972
#accession A35726
#molecule_type mRNA
#residues 1-353 #label WIL
#cross-references GB:J05262
REFERENCE A33415
#authors Sheares, B.T.; White, S.S.; Molowa, D.T.; Chan, K.; Ding,
V.D.H.; Kroon, P.A.; Bostedor, R.G.; Karkas, J.D.
#journal Biochemistry (1989) 28:8129-8135
#title Cloning, analysis, and bacterial expression of human farnesyl
pyrophosphate synthetase and its regulation in Hep G2
cells.
#cross-references MIM:90105318
#accession A33415
#molecule_type mRNA
#residues 8-115, 117-353 #label SHE
#cross-references GB:M29863

GENETICS
#gene GDB:FDPS
#cross-references GDB:128629

FUNCTION
#description catalyzes the condensation of isopentenyl diphosphate with
dimethylallyl diphosphate forming geranyl diphosphate and
pyrophosphate; catalyzes the condensation of isopentenyl
diphosphate with geranyl diphosphate forming farnesyl
diphosphate and pyrophosphate
#pathway steriodogenesis; lipoprotein synthesis
CLASSIFICATION #superfamily farnesyl-pyrophosphate synthetase
KEYWORDS cholesterol biosynthesis; homodimer; transferase
SUMMARY #length 353 #molecular-weight 40520 #checksum 7314

Query Match 48.7%; Score 55; DB 1; Length 353;
Best Local Similarity 33.3%; Pred. No. 2.89e+01;
Matches 4; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

Db 310 lpavflgyeeds 321
|:|:|:|:|:
QY 3 PPKYLHYDEET 14

Search completed: Tue Dec 2 14:52:43 1997
Job time : 12 secs.

```



\*\*\*\*\*  
W P S R E H  
\*\*\*\*\* (TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (C) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by Intelligent, Inc.

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Dec 2 14:52:04 1997; MasPar time 2.01 Seconds  
158.672 Million cell updates/sec

Similar output not generated.

File: >US-08-915-004-7  
Description: (1-15) from US08915004.pep  
Perfect Score: 113  
Sequence: 1 ETPPPKYLHYDEETS 15

Scoring table: PAM 150  
Gap 15

Searched: 59021 seqs, 21210388 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot34  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11

Statistics: Mean 27.188; Variance 37.530; scale 0.724

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	62	54.9	592	5	INVA_DAUCA BETA-FRUCTOFURANOSIDA	4.87e+01
2	60	53.1	252	9	TRYPsin IOTA PRECURSOR	1.15e+00
3	60	53.1	423	11	YPIA_BACSU HYPOTHETICAL 48.3 KD	1.15e+00
4	60	53.1	525	2	COX1_CAEEL CYTOCHROME C OXIDASE	1.15e+00
5	59	52.2	1188	3	DPO1_ADE40 DNA POLYMERASE (EC 2.	1.75e+00
6	58	51.3	643	2	CR72_BACTI 72 KD CRYSTAL PROTEIN	2.65e+00
7	56	49.6	273	11	YF30_METH HYPOTHETICAL 30.7 KD	6.01e+00
8	56	49.6	332	4	GP8D_CHLTP VIRULENCE PROTEIN PGF	6.01e+00
9	56	49.6	458	8	RBL2_RHOCA RIBULOSE BISPHOSPHATE	6.01e+00
10	56	49.6	510	11	YQBA_BACSU HYPOTHETICAL 58.5 KD	6.01e+00
11	56	49.6	525	2	COX1_ASCSU CYTOCHROME C OXIDASE	6.01e+00
12	56	49.6	678	11	YHQ_ECOLI HYPOTHETICAL 77.2 KD	6.01e+00
13	56	49.6	804	8	RRL_HSV6U RIBONUCLEOSIDE-DIPHOS	6.01e+00
14	56	49.6	1056	11	YNN2_YEAST HYPOTHETICAL 119.3 KD	6.01e+00
15	55	48.7	353	4	FPPS_HUMAN FARNESYL PYROPHOSPHAT	8.97e+00
16	55	48.7	565	6	MAOX_SCHPO MALATE OXIDOREDUCTASE	8.97e+00
17	55	48.7	586	1	ACES_TORCA ACETYLCHOLINESTERASE	8.97e+00
18	55	48.7	590	1	ACES_TORMA ACETYLCHOLINESTERASE	8.97e+00
19	55	48.7	604	7	PAPL_YEAST PUTATIVE AMINO-ACID P	8.97e+00
20	55	48.7	759	1	ARYL_CALVI ARYLPHORIN A4 PRECURS	8.97e+00
21	54	47.8	154	3	ELYS_HALWA EGG-LYSIN PRECURSOR (	1.33e+01
22	54	47.8	154	3	ELYS_HALSO EGG-LYSIN PRECURSOR (	1.33e+01

23	54	47.8	154	3	ELYS_HALRU EGG-LYSIN PRECURSOR (	1.33e+01
24	54	47.8	155	3	ELYS_HALCO EGG-LYSIN PRECURSOR (	1.33e+01
25	54	47.8	308	9	TFS2_CABEL PUTATIVE TRANSGLUTAMIN	1.33e+01
26	54	47.8	459	8	RBL2_RHOSH RIBULOSE BISPHOSPHATE	1.33e+01
27	54	47.8	466	8	RBL2_RHORU RIBULOSE BISPHOSPHATE	1.33e+01
28	54	47.8	652	5	IRGA_VIBCH IRON-REGULATED OUTER	1.33e+01
29	54	47.8	803	9	SUSY_ALNGL SUCROSE SYNTHASE (EC	1.33e+01
30	54	47.8	1036	11	YG35_YEAST HYPOTHETICAL 117.0 KD	1.33e+01
31	54	47.8	1246	11	YMW2_CABEL HYPOTHETICAL 130.8 KD	1.33e+01
32	53	46.9	190	10	VA52_VACCC PROTEIN A52.	1.97e+01
33	53	46.9	190	10	VA52_VACCC PROTEIN A52.	1.97e+01
34	53	46.9	400	6	LIPA_NEIME CAPSULE POLYSACCHARID	1.97e+01
35	53	46.9	449	4	GLVG_BACSU PROBABLE 6-PHOSPHO-BE	1.97e+01
36	53	46.9	489	1	BGAL_SULSH BETA-GALACTOSIDASE (E	1.97e+01
37	53	46.9	489	1	BGAM_SULSO BETA-GALACTOSIDASE (E	1.97e+01
38	53	46.9	491	1	BGAL_SULSO BETA-GALACTOSIDASE (E	1.97e+01
39	53	46.9	609	1	BAP2_YEAST LEU/VAL/ILE AMINO-ACI	1.97e+01
40	52	46.0	210	10	YQ06_MYCCE HYPOTHETICAL PROTEIN	2.89e+01
41	52	46.0	263	11	YIV6_YEAST HYPOTHETICAL OXIDORE	2.89e+01
42	52	46.0	470	11	YSS2_CAEEL PUTATIVE SERINE CARBO	2.89e+01
43	52	46.0	620	10	Y241_MYCCE HYPOTHETICAL PROTEIN	2.89e+01
44	52	46.0	798	10	YA73_SCHPO HYPOTHETICAL 92.1 KD	2.89e+01
45	52	46.0	1250	2	BXE_CLOBO BOTULINUM NEUROTOXIN	2.89e+01

ALIGNMENTS

RESULT	ID	INVA_DAUCA	STANDARD;	PRT;	592 AA.
AC	P26792;				
DT	01-AUG-1992 (REL. 23, CREATED)				
DT	01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)				
DT	01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)				
DE	BETA-FRUCTOFURANOSIDASE, INSOLUBLE ISOENZYME PRECURSOR (EC 3.2.1.26)				
DE	(SUCROSE-6-PHOSPHATE HYDROLASE) (INVERTASE).				
OS	DAUCUS CAROTA (CARROT).				
OC	EUKARYOTA; PLANTIA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE; APIALES; UMBELLIFERAE.				
RN	[1]				
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.				
RC	STRAIN-CV, QUEEN ANNE'S LACE;				
RX	MEDLINE; 9305650.				
RA	STURM A., CHRISPEELS M.J.;				
RL	PLANT CELL 2:1107-1119(1990).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE; 94035200.				
RA	RAMLOCH-LORENZ K., KNUDSEN S., STURM A.;				
RL	PLANT J. 4:545-554(1993).				
CC	FUNCTION: MAY PLAY AN IMPORTANT ROLE IN PHLOEM UNLOADING AND IN STRESS RESPONSE.				
CC	CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL NON-REDUCING BETA-D-FRUCTOFURANOSIDE RESIDUES IN BETA-D-FRUCTOFURANOSIDES.				
CC	SUBCELLULAR LOCATION: IONICALLY BOUND TO THE CELL WALL.				
CC	INDUCTION: IN RESPONSE TO WOUNDING AND BACTERIAL INFECTION.				
CC	SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.				
CC	THERE ARE AT LEAST THREE ISOZYMES OF BETA-FRUCTOFURANOSIDASE IN CARROT: ONE INSOLUBLE AND TWO SOLUBLE.				
CC	EMBL; M58362; G167551; -				
DR	EMBL; X69321; G18324; -				
DR	PIR; JQ0991; JQ0991.				
DR	PROSITE; PS00609; GLYCOSYL_HYDROL_F32.				
KW	HYDROLASE; GLYCOSIDASE; GLYCOPROTEIN; CELL WALL; ZYMOGEN; SIGNAL.				
FT	SIGNAL 1 31				
FT	PROPEP 32 48				
FT	CHAIN 49 592				
FT	ACT_SITE 74 74				
FT	CARBOHYD 170 170				
FT	CARBOHYD 195 195				
FT	CARBOHYD 311 311				
FT	CARBOHYD 348 348				
FT	CARBOHYD 570 570				

FT CONFLICT 133 133 R -> W (IN REF. 2).  
FT CONFLICT 487 487 A -> V (IN REF. 2).  
SQ SEQUENCE 592 AA; 66813 MW; D41B4A3C CRC32;

Query Match 54.9%; Score 62; DB 5; Length 592;  
Best Local Similarity 61.5%; Pred. No. 4.87e+01;  
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 435 epfdpkwleydae 447  
QY 1 ETPPKYLHYDEE 13

RESULT 2  
ID TRYI\_DROME STANDARD; PRT; 252 AA.  
AC P52905;  
DT 01-OCT-1996 (REL. 34, CREATED)  
01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
TRYPSIN IOTA PRECURSOR (EC 3.4.21.4).  
GN TRY-TOTA.  
OS DROSOPHILA MELANOGASTER (FRUIT FLY).  
OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-OREGON-R;  
RA GAO L., WANG S., HICKEY D.A.;  
RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-|-XAA, LYS-|-XAA.  
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY.  
EMBL; U41476; G1134972; -.  
DR FLYBASE; FBN0015001; TRY-TOTA.  
KW HYDROLASE; SERINE PROTEASE; ZWMOGEN; SIGNAL; MULTIGENE FAMILY.  
FT SIGNAL 1 19  
FT PROPEP 20 27  
FT CHAIN 28 252  
FT ACT\_SITE 68 68  
FT ACT\_SITE 113 113  
FT ACT\_SITE 206 206  
FT DISULFID 53 69  
FT DISULFID 175 193  
FT DISULFID 202 226  
FT SITE 200 200  
SQ SEQUENCE 252 AA; 26612 MW; CE103ACF CRC32;

Query Match 53.1%; Score 60; DB 9; Length 252;  
Best Local Similarity 54.5%; Pred. No. 1.15e+00;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 103 eqfdrflhyd 113  
QY 1 ETPPKYLHYD 11

RESULT 3  
ID YPIA\_BACSU STANDARD; PRT; 423 AA.  
AC P54389;  
DT 01-OCT-1996 (REL. 34, CREATED)  
01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 48.3 KD PROTEIN IN OCRA 5'REGION.  
GN YPIA.  
OS BACILLUS SUBTILIS.  
OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-166 / MAREBURG;  
RA SOROKIN A.V., AZEVEDO V., ZUMSTEIN E., GALLERON N., EHRlich S.D.,  
RA SERROR P.;  
RL SUBMITTED (JAN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; L47709; G1146224; -.

DR SUBTILIST; BG11496; YPIA.  
KW HYPOTHETICAL PROTEIN.  
SQ SEQUENCE 423 AA; 48321 MW; 708A0553 CRC32;

Query Match 53.1%; Score 60; DB 11; Length 423;  
Best Local Similarity 46.2%; Pred. No. 1.15e+00;  
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 361 qsfcaaylhyred 373  
QY 1 ETPPKYLHYDEE 13

RESULT 4  
ID COXL\_CAEEL STANDARD; PRT; 525 AA.  
AC P24893;  
DT 01-MAR-1992 (REL. 21, CREATED)  
01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).  
GN COI.  
OS CAENORHABDITIS ELEGANS.  
OC MITOCHONDRION.  
OC EUKARYOTA; METAZOA; ACLOLOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE; 92201635.  
RA OKIMOTO R., MACFARLANE J.L., CLARY D.O., WOLSTENHOLME D.R.;  
RL GENETICS 130:471-498 (1992).  
CC -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2 AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3 AND COPPER B.  
CC -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.  
CC -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O + 4 FERRICYTOCHROME C.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.  
CC -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.  
DR EMBL; X54252; G515883; -.  
DR PIR; S26034; S26034.  
DR PROSITE; PS00077; COX1  
KW OXIDOREDUCTASE; HEME; COPPER; MITOCHONDRION; TRANSMEMBRANE;  
RESPIRATORY CHAIN; INNER MEMBRANE.  
FT METAL 69 69  
FT METAL 247 247  
FT METAL 251 251  
FT METAL 297 297  
FT METAL 298 298  
FT METAL 383 383  
FT METAL 385 385  
SQ SEQUENCE 525 AA; 58463 MW; FBD63AAD CRC32;

Query Match 53.1%; Score 60; DB 2; Length 525;  
Best Local Similarity 60.0%; Pred. No. 1.15e+00;  
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 443 fprkyldypd 452  
QY 3 ETPPKYLHYDE 12

RESULT 5  
ID DPOL\_ADE40 STANDARD; PRT; 1188 AA.  
AC P48311;  
DT 01-FEB-1996 (REL. 33, CREATED)  
01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
DE DNA POLYMERASE (EC 2.7.7.7).

GN POL.  
 OS HUMAN ADENOVIRUS TYPE 40.  
 OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; ADENOVIRIDAE; MASTADENOVIRUSES.  
 RN [1]  
 RC STRAIN-MARBURG;  
 RX MEDLINE; 90221821.  
 RA BOKRANZ M., KLEIN A., MEILE L.;  
 RC STRAIN-DUGAN;  
 RX MEDLINE; 94087748.  
 RA DAVISON A.J., TELFORD E.A., WATSON M.S., MCBRIDE K., MAUTNER V.;  
 RL J. MOL. BIOL. 234:1308-1316(1993).  
 CC -!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE =  
 CC N PYROPHOSPHATE + DNA(N).  
 CC -!- THIS DNA POLYMERASE REQUIRES A PROTEIN AS A PRIMER.  
 CC -!- SIMILARITY: BELONGS TO FAMILY B OF DNA POLYMERASES.  
 DR EMBL; L19443; G303975; -.  
 KW DNA-DIRECTED DNA POLYMERASE; DNA REPLICATION; DNA-BINDING.  
 SQ SEQUENCE 1188 AA; 135289 MW; C8E56098 CRC32;

Query Match 52.2%; Score 59; DB 3; Length 1188;  
 Best Local Similarity 60.0%; Pred. No. 1.75e+00;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

377 fspdylytyee 386

1:1 1:1 1:1  
 3 FPKYLHYDE 12

RESULT 6  
 ID CR72\_BACTI STANDARD; PRT; 643 AA.  
 AC P21256; 01-MAY-1991 (REL. 18, CREATED)  
 DT 01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)  
 DT 01-MAY-1991 (REL. 18, LAST ANNOTATION UPDATE)  
 DE 72 KD CRYSTAL PROTEIN (DELTA ENDOTOXIN) (CRYSTALLINE ENTOMOCIDAL  
 DE PROTOXIN).  
 GN CRYD.  
 OS BACILLUS THURINGIENSIS (SUBSP. ISRAELENSIS).  
 OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 89008093.  
 RA DONOVAN W.P., DANKOCISK C., GILBERT M.P.;  
 RL J. BACTERIOL. 170:4732-4738(1988).  
 CC -!- FUNCTION: PROMOTES COLLOIDSMOTIC LYSIS BY BINDING TO THE MIDGUT  
 CC EPITHELIAL CELLS OF MOSQUITOS.  
 CC -!- TOXIC SEGMENT OF THE PROTEIN IS LOCATED IN THE N-TERMINUS.  
 CC -!- DEVELOPMENTAL STAGE: THE CRYSTAL PROTEIN IS PRODUCED DURING  
 CC SPORE-GERMINATION AND IS ACCUMULATED BOTH AS AN INCLUSION AND AS PART  
 CC OF THE SPORE COAT.  
 DR EMBL; M31737; G142763; -.  
 PIR; A43647; A43647.  
 TOXIN; SPORE-GERMINATION.  
 SQ SEQUENCE 643 AA; 72348 MW; 0C528C2C CRC32;

Query Match 51.3%; Score 58; DB 2; Length 643;  
 Best Local Similarity 70.0%; Pred. No. 2.65e+00;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 413 tippnylyad 422  
 1:1 1:1 1:1  
 2 TFPKYLHYD 11

RESULT 7  
 ID YP30\_METH STANDARD; PRT; 273 AA.  
 AC P14932;  
 DT 01-APR-1990 (REL. 14, CREATED)  
 DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)  
 DT 01-APR-1990 (REL. 14, LAST ANNOTATION UPDATE)  
 DE HYPOTHETICAL 30.7 KD PROTEIN.  
 OS METHANOBACTERIUM THERMAUTOTROPHICUM.  
 OC PLASMID PME2001.  
 OC ARCHAEABACTERIA; EURYARCHAEOTA; METHANOBACTERIALES;  
 OC METHANOBACTERIACEAE.

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MARBURG;  
 RX MEDLINE; 90221821.  
 RA BOKRANZ M., KLEIN A., MEILE L.;  
 RL NUCLEIC ACIDS RES. 18:363-363(1990).  
 DR EMBL; X17205; G45718; -.  
 PIR; S07880; S07880.  
 KW HYPOTHETICAL PROTEIN; PLASMID.  
 SQ SEQUENCE 273 AA; 30744 MW; FF211F2C CRC32;

Query Match 49.6%; Score 56; DB 11; Length 273;  
 Best Local Similarity 58.3%; Pred. No. 6.01e+00;  
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 223 spkylfysatg 234  
 1:1 1:1 1:1  
 4 PPKYLHYDETS 15

RESULT 8  
 ID GP8D\_CHLTR STANDARD; PRT; 332 AA.  
 AC P08788; P10554;  
 DT 01-NOV-1988 (REL. 09, CREATED)  
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
 DE VIRULENCE PROTEIN PGP8-D (ORF8) (PROTEIN N-1/N-2).  
 OS CHLAMYDIA TRACHOMATIS.  
 OC PLASMID PLGV440, AND PLASMID PCHL1.  
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; RICKETTSIAS; CHLAMYDIALES;  
 OC CHLAMYDIACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX PLASMID-PLGV440; STRAIN-L2;  
 RC MEDLINE; 89013895.  
 RA COMANDUCCI M., RICCI S., RATTI G.;  
 RL MOL. MICROBIOL. 2:531-538(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX PLASMID-PLGV440; STRAIN-L1/440/LN;  
 RC MEDLINE; 89233998.  
 RA HAIT C., WARD M.E., CLARKE I.N.;  
 RL NUCLEIC ACIDS RES. 16:4053-4067(1988).  
 RN [3]  
 RP REVISIONS TO C-TERMINUS.  
 RC PLASMID-PLGV440; STRAIN-L1/440/LN;  
 RA HAIT C.;  
 RL SUBMITTED (APR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.

[4]  
 RN SEQUENCE FROM N.A.  
 RP PLASMID-PCHL1; STRAIN-TRACHOMA SEROTYPE D;  
 RC MEDLINE; 90301796.  
 RA COMANDUCCI M., RICCI S., CEVENINI R., RATTI G.;  
 RL PLASMID 23:149-154(1990).  
 CC -!- PGP8-D IS REQUIRED FOR GROWTH WITHIN MAMMALIAN CELLS.  
 CC EMBL; X07547; G40731; -.  
 DR EMBL; J03321; G455146; -.  
 DR PIR; S01180; S01180.  
 DR PIR; S01181; S01181.  
 DR PIR; S01920; S01920.  
 DR PIR; H37386; H37386.  
 KW PLASMID.  
 FT CONFLICT  
 SQ SEQUENCE 332 AA; 37918 MW; 1DEB5F46 CRC32;

Query Match 49.6%; Score 56; DB 4; Length 332;  
 Best Local Similarity 41.7%; Pred. No. 6.01e+00;  
 Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 225 typqkfmhflqe 236  
 1:1 1:1 1:1  
 2 TFPKYLHYDE 13

RESULT 9  
ID RBL2LHQA STANDARD; PRT; 458 AA.  
AC P50922;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE RIBULOSE BIPHOSPHATE CARBOXYLASE LARGE CHAIN 2 (EC 4.1.1.39).  
GN CBM OR RBLP.  
OS RHODORHIZOPHYTES (RHODOSPHEUDOMONAS CAPSULATA).  
OC PROKARYOTA; GRACILICUTES; ANOXYPHOTOBACTERIA; PURPLE BACTERIA;  
OC RHODOSPIRILLACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 11166;  
RA LARIMER F.W., LU T.Y.S., BULEY D.M.;  
SUBMITTED (APR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
-1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF  
D-RIBULOSE 1,5-BIPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC  
CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF  
THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH  
REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME  
ACTIVE SITE.  
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BIPHOSPHATE + CO(2) -  
CC 2 3-PHOSPHO-D-GLYCERATE.  
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BIPHOSPHATE + O(2) -  
CC 3-PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.  
CC -1- SUBUNIT: IN CONTRAST TO FORM I GENES, THE FORM II RUBISCO ARE  
CC COMPOSED OF ONLY LARGE SUBUNITS (BY SIMILARITY).  
CC -1- THIS PROTEIN IS ENCODED WITHIN THE FORM II RIBULOSE-BIPHOSPHATE  
CC CARBOXYLASE OPERON.  
DR EMBL: U23145; G727418; -  
KW PHOTOSYNTHESIS; CARBON DIOXIDE FIXATION; PHOTORESPIRATION;  
KW LPASE; OXIDOREDUCTASE; MONOOXYGENASE; MULTIGENE FAMILY.  
FT ACT SITE 191 191 BINDING OF CO(2) ACTIVATES THE ENZYME.  
SQ SEQUENCE 458 AA; 50091 MW; EAAE2535 CRC32;  
Query Match 49.6%; Score 56; DB 8; Length 458;  
Best Local Similarity 55.0%; Pred. No. 6.01e+00;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
Db 278 sfrpqfhy 286  
QY 2 fppkylhy 10  
RESULT 10  
ID YQBA\_BACSU STANDARD; PRT; 510 AA.  
AC P45917;  
DT 01-NOV-1995 (REL. 32, CREATED)  
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 58.5 KD PROTEIN IN SPOILIC-CWLA INTERGENIC REGION.  
GN YQBA.  
OS BACILLUS SUBTILIS.  
OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-168 / JH642;  
RC MEDLINE: 95219086;  
RA TAKEMARU K.-I., MIZUNO M., SATO T., TAKEUCHI M., KOBAYASHI Y.;  
RL MICROBIOLOGY 141:323-327(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-168 / JH642;  
RA KOBAYASHI Y., MIZUNO M., MASUDA S., TAKEMARU K., HOSONO S., SATO T.,  
RA TAKEUCHI M.;  
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RP IDENTIFICATION.  
RX MEDLINE: 96084975.  
RA MEDIGUE C., MOSZER I., VIARI A., DANCHIN A.;

RL GENE 165:GC37-GC51(1995).  
CC -1- SIMILARITY: STRONG, TO B.SUBTILIS XKDE.  
DR EMBL: D32216; G1217849; -  
DR EMBL: D84432; G1303740; -  
DR SUBTILIST; BGL1272; YQBA.  
KW HYPOTHETICAL PROTEIN.  
SQ SEQUENCE 510 AA; 58458 MW; 433165AB CRC32;  
Query Match 49.6%; Score 56; DB 11; Length 510;  
Best Local Similarity 56.7%; Pred. No. 6.01e+00;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
Db 105 kylhdesa 113  
QY 6 KYLHYDEET 14  
RESULT 11  
ID COX1LASCUS STANDARD; PRT; 525 AA.  
AC P24881;  
DT 01-MAR-1992 (REL. 21, CREATED)  
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).  
GN COI.  
OS ASCARIS SUUM (PIG ROUNDWORM) (ASCARIS LUMBRICOIDES).  
OG MITOCHONDRION.  
OC EUCARYOTA; METAZOA; ACCELOMATES; NEMATODA; SECERNENTEA; ASCARIDIDA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BODY WALL MUSCLE, AND EGG;  
RX MEDLINE: 92201635.  
RA OKIMOTO R., MACFARLANE J.L., CLARY D.O., WOLSTENHOLME D.R.;  
RL GENETICS 130:471-498(1992).  
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-  
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE  
CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN  
CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2  
CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3  
CC AND COPPER B.  
CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.  
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) - 2 H(2)O +  
CC 4 FERROCYTOCHROME C.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL  
CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.  
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.  
DR EMBL: X54253; G559496; -  
DR PIR: S26022; S26022.  
DR PROSITE: PS00077; COX1.  
KW OXIDOREDUCTASE; HEME; COPPER; MITOCHONDRION; TRANSMEMBRANE;  
KW RESPIRATORY CHAIN; INNER MEMBRANE.  
FT METAL 69 69 IRON (HEME A) (PROBABLE).  
FT METAL 247 247 COPPER B (PROBABLE).  
FT METAL 251 251 COPPER B (PROBABLE).  
FT METAL 297 297 COPPER B (PROBABLE).  
FT METAL 298 298 COPPER B (PROBABLE).  
FT METAL 383 383 IRON (HEME A3) (PROBABLE).  
FT METAL 385 385 IRON (HEME A) (PROBABLE).  
SQ SEQUENCE 525 AA; 58529 MW; A6124BE2 CRC32;  
Query Match 49.6%; Score 56; DB 2; Length 525;  
Best Local Similarity 50.0%; Pred. No. 6.01e+00;  
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
Db 443 yprkylhydp 452  
QY 3 FPPKYLHYDE 12  
RESULT 12  
ID YHQ\_ECOLI STANDARD; PRT; 678 AA.  
AC P32138;

DT 01-OCT-1993 (REL. 27, CREATED)  
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)  
DT 01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE)  
DE HYPOHETICAL 77.2 KD PROTEIN IN GLNA-FDHE INTERGENIC REGION (F678).  
GN YIHO.  
OS ESCHERICHIA COLI.  
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
OC ENTEROBACTERIACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / MG1655;  
RX MEDLINE: 93347969.  
RA PLUNKETT G. III, BURLAND V.D., DANIELS D.L., BLATTNER F.R.;  
RL NUCLEIC ACIDS RES. 21:3391-3398(1993).  
DR EMBL: L19201; G304982; -;  
DR PIR: S40822; S40822.  
DR ECGENE; EG11843; YIHO.  
KW HYPOHETICAL PROTEIN.  
SQ SEQUENCE 678 AA; 77249 MW; 8D74E4AE CRC32;  
Query Match 49.6%; Score 56; DB 11; Length 678;  
Best Local Similarity 40.0%; Pred. No. 6.01e+00;  
Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;  
Db 589 plflhyedda 598  
| :|||::: 14  
QY 5 PKYLHYDE 14  
RESULT 13  
ID R1RL\_HSV6U STANDARD; PRT; 804 AA.  
AC P52343;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE LARGE CHAIN (EC 1.17.4.1)  
DE (RIBONUCLEOTIDE REDUCTASE).  
GN U28 OR P2LF2  
OS HERPES SIMPLEX VIRUS (TYPE 6 / STRAIN UGANDA-1102).  
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; BETAHERPESVIRINAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 95266321.  
RA GOMPELS U.A., NICHOLAS J., LAWRENCE G., JONES M., THOMSON B.J.,  
RA MARTIN M.E., EFSTATHIOU S., CRAXTON M., MACAULAY H.A.;  
RL VIROLOGY 209:29-51(1995).  
RC -!- FUNCTION: PROVIDES THE PRECURSORS NECESSARY FOR DNA SYNTHESIS.  
CC -!- CATALYTIC ACTIVITY: 2'-DEOXYRIBONUCLEOSIDE DIPHOSPHATE + OXIDIZED  
CC THIOREDUXIN + H(2)O = RIBONUCLEOSIDE DIPHOSPHATE + REDUCED  
CC THIOREDUXIN.  
-!- PATHWAY: FIRST REACTION IN THE DNA REPLICATION PATHWAY.  
-!- SUBUNIT: HETERODIMER OF A LARGE AND A SMALL CHAIN.  
CC -!- SIMILARITY: HIGH TO OTHER EUKARYOTIC, PROKARYOTIC, AND VIRAL  
CC RIBONUCLEOSIDE DIPHOSPHATE REDUCTASE LARGE CHAINS.  
DR EMBL: X83413; G854007; -;  
KW OXIDOREDUCTASE; DNA REPLICATION.  
SQ SEQUENCE 804 AA; 93352 MW; B599A3FD CRC32;  
Query Match 49.6%; Score 56; DB 8; Length 804;  
Best Local Similarity 50.0%; Pred. No. 6.01e+00;  
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
Db 365 tftskylefe 374  
| :|||::: 11  
QY 2 TFPKYLHYD 11  
RESULT 14  
ID YNN2\_YEAST STANDARD; PRT; 1056 AA.  
AC P53914;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)

DE HYPOHETICAL 119.3 KD PROTEIN IN FPRI-TOM22 INTERGENIC REGION.  
GN YNL132W OR N1216 OR N1858.  
OS SACHAROMYCES CEREVISIAE (BAKER'S YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.,  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-S288C;  
RX MEDLINE: 96109932.  
RA MALLET L., BUSSEAU F., JACQUET M.;  
RL YEAST 11:1195-1209(1995).  
CC -!- SIMILARITY: TO AN A.AMBISEXUALIS HYPOHETICAL PROTEIN (AC P54008).  
DR EMBL: 246843; G854505; -;  
DR EMBL: 271408; E239944; -;  
KW HYPOHETICAL PROTEIN; ATP-BINDING  
FT NP\_BIND 283 290  
SQ SEQUENCE 1056 AA; 119347 MW; CF839C98 CRC32;  
Query Match 49.6%; Score 56; DB 11; Length 1056;  
Best Local Similarity 85.7%; Pred. No. 6.01e+00;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Db 716 pphyhy 722  
| :|||::: 10  
QY 4 PPKYLHY 10  
RESULT 15  
ID FPPS\_HUMAN STANDARD; PRT; 353 AA.  
AC P14324;  
DT 01-JAN-1990 (REL. 13, CREATED)  
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
DE FARNESYL PYROPHOSPHATE SYNTHETASE (FPP SYNTHETASE) (FPS) (FARNESYL  
DE DIPHOSPHATE SYNTHETASE) (DIMETHYLLALLYLTRANSFERASE (EC 2.5.1.1) /  
DE GERANYLTRANSFERASE (EC 2.5.1.10)) (KIRA0032).  
GN FDPFS OR FPS.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 90170972.  
RA WILKIN D.J., KUTSUNAI S.Y., EDWARDS P.A.;  
RL J. BIOL. CHEM. 265:4607-4614(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA NOMURA N., MIYAJIMA N., KANARABAYASHI Y., TABATA S.;  
RL SUBMITTED (MAR-1993) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RP SEQUENCE OF 8-353 FROM N.A.  
RC TISSUE=LIVER;  
RX MEDLINE: 90105318.  
RA SHEARES B.T., WHITE S.S., MOLOWA D.T., CHAN K., DING V.D.-H.,  
RA KROON P.A., BOSTEDOR R.G., KARKAS J.D.;  
RL BIOCHEMISTRY 28:8129-8135(1989).  
CC -!- FUNCTION: CATALYZES THE SEQUENTIAL CONDENSATION OF ISOPENTENYL  
CC PYROPHOSPHATE WITH THE ALLYLIC PYROPHOSPHATES, DIMETHYLLALLYL  
CC PYROPHOSPHATE, AND THEN WITH THE RESULTANT GERANYLPYROPHOSPHATE  
CC TO THE ULTIMATE PRODUCT FARNESYL PYROPHOSPHATE.  
CC -!- CATALYTIC ACTIVITY: DIMETHYLLALLYL DIPHOSPHATE + ISOPENTENYL  
CC DIPHOSPHATE = PYROPHOSPHATE + GERANYL DIPHOSPHATE.  
CC -!- CATALYTIC ACTIVITY: GERANYL DIPHOSPHATE + ISOPENTENYL DIPHOSPHATE  
CC = PYROPHOSPHATE + TRANS,TRANS-FARNESYL DIPHOSPHATE.  
CC -!- SUBUNIT: HOMODIMER.  
CC -!- PATHWAY: ISOPRENE BIOSYNTHESIS, CHOLESTEROL BIOSYNTHESIS.  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -!- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS TOGETHER FPP  
CC SYNTHETASE, GGPP SYNTHETASE AND HEXAPRENIL PYROPHOSPHATE  
CC SYNTHETASE.  
DR EMBL: J05262; G182399; -;  
DR EMBL: D14697; G285965; ALT\_INIT.  
DR EMBL: M29863; G182405; -;  
DR PIR: A35726; A35726.

DR MIM; 134631; -.  
DR PROSITE; PS00444; POLYPRENYL\_SYNTHET\_2.  
DR PROSITE; PS00723; POLYPRENYL\_SYNTHET\_1.  
KW TRANSFERASE; ISOPRENE BIOSYNTHESIS; CHOLESTEROL BIOSYNTHESIS.  
FT ACT\_SITE 192 192 BY SIMILARITY.  
FT CONFLICT 116 116 I -> T (IN REF. 1).  
SQ SEQUENCE 353 AA; 40532 MW; 87C45ECD CRC32;

Query Match 48.7%; Score 55; DB 4; Length 353;  
Best Local Similarity 33.3%; Pred. No. 8.97e-00;  
Matches 4; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

Db 310 lpavilqyeeds 321  
QY 3 FPPKYLHYDEET 14

Search completed: Tue Dec 2 14:52:14 1997  
time : 10 secs.



\*\*\*\*\*

WPIREH (TM)

\*\*\*\*\*

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

MPSrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 2 14:43:20 1997; Maspar time 990.46 Seconds  
1313.995 Million cell updates/sec  
\*\*\*\*\*  
ular output not generated.

Title: >US-08-915-004-8  
Description: (11185) from US08915004.seq  
Perfect Score: 1185  
N.A. Sequence: 1 ATGAACAACCTGCTGCTG.....TAAATAAGCTGCTATAA 1185  
Comp: TACTTGTGAACGACGAC.....ATTTTATTCGACGATATT

Scoring table:

Gap 6

Nmatch STD : Dbase 0; Query 0

Searches: 362067 seqs, 549138275 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database:

1: BCT 2: FUN 3: GEN1 4: GEN2 5: HTG1 6: HTG2 7: HUM 8: INV  
9: OCT 10: MAM 11: VRT 12: PLN 13: PRO 14: ROD 15: SYN 16: UNC  
17: VTR

Database:

genbank99  
18: BCT1 19: BCT2 20: BCT3 21: BCT4 22: BCT5 23: BCT6 24: BCT7  
25: BCT8 26: BCT9 27: BCT10 28: BCT11 29: GEN1 30: GEN2  
31: GEN3 32: HTG1 33: HTG2 34: HTG3 35: INV1 36: INV2 37: INV3  
38: INV4 39: INV5 40: INV6 41: INV7 42: INV8 43: INV9 44: INV10  
45: INV11 46: MAM1 47: MAM2 48: MAM3 49: VRT1 50: VRT2 51: VRT3  
52: VRT4 53: PAT1 54: PAT2 55: PAT3 56: PAT4 57: PAT5 58: PHG  
59: PLN1 60: PLN2 61: PLN3 62: PLN4 63: PLN5 64: PLN6 65: PLN7  
66: PLN8 67: PLN9 68: PLN10 69: PLN11 70: PR11 71: PR12  
72: PR13 73: PR14 74: PR15 75: PR16 76: PR17 77: PR18 78: PR19  
79: PR20 80: PR21 81: PR22 82: PR23 83: PR24 84: PR25  
85: ROD1 86: ROD2 87: ROD3 88: ROD4 89: ROD5 90: ROD6 91: ROD7  
92: ROD8 93: STR 94: SYN 95: UNA 96: VRL1 97: VRL2 98: VRL3  
99: VRL4 100: VRL5 101: VRL6 102: VRL7 103: VRL8 104: VRL9  
105: VRL10  
genbank-new3  
106: BCT 107: GEN1 108: GEN2 109: HTG1 110: HTG2 111: INV  
112: MAM 113: VRT 114: PHG 115: PLN 116: PR11 117: PR12  
118: ROD 119: SYN 120: UNA 121: VRL  
u-emb150\_99  
122: part1

Database:

Statistics: Mean 11.245; Variance 4.362; scale 2.578

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
C 1	32	2.7	215	57	Sequence 5 from paten	2.40e-06
C 2	30	2.5	215	57	Sequence 5 from paten	7.05e-05
C 3	28	2.4	354	111	Oxytricha fallax 57kD	1.84e-03
C 4	28	2.4	354	8	Oxytricha fallax 57kD	1.84e-03
C 5	25	2.1	354	111	Oxytricha fallax 57kD	1.90e-01
C 6	25	2.1	354	8	Oxytricha fallax 57kD	1.90e-01
C 7	25	2.1	1441	51	glial fibrillary acid	1.90e-01
C 8	25	2.1	1933	49	Carassius auratus (GF	1.90e-01
C 9	24	2.0	115419	34	Human DNA sequence **	8.23e-01
C 10	22	1.9	370	8	Oxytricha fallax 57kD	1.34e-01
C 11	22	1.9	370	111	Oxytricha fallax 57kD	1.34e-01
C 12	22	1.9	1256	47	O. coniculus SP17 gene	1.34e-01
C 13	22	1.9	1380	91	R. norvegicus interieu	1.34e-01
C 14	22	1.9	1388	86	M. musculus tumor necr	1.34e-01
C 15	22	1.9	1388	118	M. musculus tumor necr	1.34e-01
C 16	22	1.9	1505	89	Mouse tumor necrosis	1.34e-01
C 17	22	1.9	1761	67	S. cerevisiae chromoso	1.34e-01
C 18	23	1.9	1851	117	H. sapiens LIPA gene,	3.40e+00
C 19	23	1.9	1851	75	H. sapiens LIPA gene,	3.40e+00
C 20	22	1.9	2492	65	S. cerevisiae fus2 gen	1.34e-01
C 21	22	1.9	2552	55	Sequence 5 from Paten	1.34e-01
C 22	22	1.9	2787	91	Rat peroxisomal 3-ket	1.34e-01
C 23	23	1.9	3021	69	S. cerevisiae nitrogen	3.40e+00
C 24	22	1.9	3112	81	Human LD78 beta gene.	1.34e-01
C 25	23	1.9	3177	43	L. stagnalis mRNA for	3.40e+00
C 26	22	1.9	3699	67	S. cerevisiae chromoso	1.34e-01
C 27	22	1.9	3796	90	Murine tumor necrosis	1.34e-01
C 28	22	1.9	4561	38	Caenorhabditis elegan	1.34e-01
C 29	22	1.9	7057	88	Mouse mRNA for mouse	1.34e-01
C 30	22	1.9	17013	20	E. amylovora (Ea7/74)	1.34e-01
C 31	22	1.9	26923	38	Caenorhabditis elegan	1.34e-01
C 32	22	1.9	28687	37	C. elegans cosmid F42	1.34e-01
C 33	23	1.9	38586	37	Caenorhabditis elegan	3.40e+00
C 34	22	1.9	39307	65	S. cerevisiae DNA for	1.34e-01
C 35	22	1.9	40367	77	Human DNA sequence fr	1.34e-01
C 36	22	1.9	40397	65	S. cerevisiae chromoso	1.34e-01
C 37	23	1.9	43100	65	Saccharomyces cerevis	3.40e+00
C 38	22	1.9	90468	34	Human DNA sequence **	1.34e-01
C 39	23	1.9	145750	33	Human DNA sequence **	3.40e+00
C 40	22	1.9	185775	80	Homo sapiens fragile	1.34e-01
C 41	22	1.9	257912	103	Paramedium bursaria C	1.34e-01
C 42	21	1.8	105	56	Sequence 13 from pate	5.01e-01
C 43	21	1.8	37844	110	Caenorhabditis elegan	5.01e-01
C 44	21	1.8	112762	117	Human DNA sequence fr	5.01e-01
C 45	21	1.8	201239	109	Human DNA sequence **	5.01e-01

ALIGNMENTS

RESULT 1 128278 215 bp DNA PAT 30-OCT-1996  
LOCUS Sequence 5 from patent US 5569830.  
DEFINITION 128278  
ACCESSION g1819054  
NID  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 215)  
AUTHORS Bennett, A., Labavitch, J. M., Powell, A. and Stotz, H.  
TITLE Plant inhibitors of fungal polygalacturonases and their use to control fungal disease  
JOURNAL Patent: US 5569830-A 5 29-OCT-1996;

FEATURES  
Location/Qualifiers  
1..215  
/organism="unknown"

BASE COUNT 15 a 8 c 25 g 26 t 141 others

ORIGIN



Q







JOURNAL Submitted (26-NOV-1993) E.E. Powell, c/o John Todd, Level 6  
Nuffield Dept of Surgery, The John Radcliffe Hospital, Oxford, UK  
REFERENCE 2 (bases 1 to 1388)  
AUTHORS Powell,E.E., Wicker,L.S., Peterson,L.B. and Todd,J.A.  
TITLE Amino acid variation in the tumor Necrosis factor receptor 2 is  
linked to autoimmune diabetes in NOD mice  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 1388)  
AUTHORS Powell,E.E., Wicker,L.S., Peterson,L.B. and Todd,J.A.  
TITLE Allelic variation of the type 2 tumor necrosis factor receptor gene  
JOURNAL Mamm. Genome 5 (11), 726-727 (1994)  
MEDLINE 95178848  
FEATURES Location/Qualifiers  
source 1..1388  
/organism="Mus musculus"  
/strain="NOD"  
/chromosome="4 (distal region)"  
CDS <1..1380  
/gene="murine tumour necrosis factor receptor 2"  
/codon\_start=1  
/db\_xref="PID:g433831"  
/translation="LWATGTPVPAQVLTTPKPEGYEQISQIYRKAQMCCAKCP  
PGQVYKFCNKTSIDTVCADCEASMTQVMNQFTCLSCSSSCSTDQVETRACTKQNR  
VCACEAGRYCALTKTSGSCRCQCMRLSKCGPGGVASSRPNGLVCKACAPGTFSDT  
SSTDVCPRHICILAIIPGNASTDAVCAPESTLSAIPRTILYVQPEPTSQBLDQEP  
GPSTQPSILTSLGSDTPIIIOSTGGISGLIGLIVGTSLGLMLGLVNCFLIVORKKK  
PSCIQROAKYPHYPDKSDQADVLGEQHLHTTAPSSSSLESASAGDRRAPPGHG  
QADYMAQSQSQARASRRISDSHSGSHGTHVNTVCIVNVCSSSDHSSQSSQASATV  
GDYDAKSASQPKQEPFQSECPSPQYETTETTLQSHKEKPLPLGVDMGMPQSQAG  
FDQLAVKVA"  
variation replace(260,"c")  
/gene="murine tumour necrosis factor receptor 2"  
variation replace("Ser to Thr"  
replace(278,"t")  
/gene="murine tumour necrosis factor receptor 2"  
variation /note="Thr to Ile"  
replace(489,"t")  
/gene="murine tumour necrosis factor receptor 2"  
/note="silent"  
variation replace(802,"a")  
/gene="murine tumour necrosis factor receptor 2"  
variation /note="Phe to Ile"  
replace(921,"c")  
/gene="murine tumour necrosis factor receptor 2"  
/note="silent"  
variation replace(1034,"t")  
/gene="murine tumour necrosis factor receptor 2"  
/note="Ser to Phe"  
variation replace(1047,"t")  
/gene="murine tumour necrosis factor receptor 2"  
/note="silent"  
variation replace(1143,"t")  
/gene="murine tumour necrosis factor receptor 2"  
/note="silent"  
variation replace(1173,"a")  
/gene="murine tumour necrosis factor receptor 2"  
/note="silent"  
variation replace(1262,"g")  
/gene="murine tumour necrosis factor receptor 2"  
/note="Thr to Cys"  
variation replace(1317,"g")  
/gene="murine tumour necrosis factor receptor 2"  
/note="silent"

```
variation      975 /gene="murine tumour necrosis factor receptor 2"
                /note="silent"
                /replace="c"
variation      1034 /gene="murine tumour necrosis factor receptor 2"
                /note="Ser to Phe"
                /replace="t"
variation      1047 /gene="murine tumour necrosis factor receptor 2"
                /note="silent"
                /replace="t"
variation      1143 /gene="murine tumour necrosis factor receptor 2"
                /note="silent"
                /replace="t"
variation      1173 /gene="murine tumour necrosis factor receptor 2"
                /note="silent"
                /replace="a"
variation      1262 /gene="murine tumour necrosis factor receptor 2"
                /note="Thr to Cys"
                /replace="g"
variation      1317 /gene="murine tumour necrosis factor receptor 2"
                /note="silent"
                /replace="g"

BASE COUNT    327 a 417 c 371 g 273 t
ORIGIN

Query Match    1.9%; Score 22; DB 118; Length 1388;
Best Local Similarity 64.1%; Pred. No. 1.34e+01;
Matches 50; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

Db 108 tcagatgtcgtgtcctaagtctcctcctggccaatatgtgaacatttctgcaacaagac 167
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 111 TCAGCTGTGTGTGACAAATGTCTCTCTGTTACCTCTAAACACACACTGTACAGCAA 170
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 168 ctcggacacccgtgtgtgc 185
      | | ||||| |||
Qy 171 GTGGAAGACCGTGTGCGC 188
```

Search completed: Tue Dec 2 15:19:52 1997  
Job time : 2192 secs.



\*\*\*\*\*  
 W P E R E H  
 \*\*\*\*\*  
 (TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
 Distribution rights by Intelligenetics, Inc.

MPSrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 2 15:20:11 1997; MasPar time 129.65 Seconds  
 953.937 Million cell updates/sec

Linear output not generated.

Title: >US-08-915-004-8  
 Description: (1-1185) from US08915004.seq  
 Perfect Score: 1185  
 N.A. Sequence: 1 ATGACACACTGCTGCTG.....TAAATATAGCTGCTTATAA 1185  
 Comp: TACTTGTGAACGACGACGAC.....ATTITATTCGACGATAAT

Scoring table: TABLE default  
 Gap 6

Mismatch STD : Dbase 0; Query 0

Searched: 142080 seqs, 52183452 bases x 2

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: n-geneSeq28  
 1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
 14:part14 15:part15 16:part16 17:part17 18:part18  
 19:part19 20:part20 21:part21 22:part22 23:part23  
 24:part24 25:part25 26:part26 27:part27 28:part28  
 29:part29

Statistics: Mean 9.128; Variance 5.035; scale 1.813

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	1059	89.4	1206	28	T36685 Osteoclastogenesis in	0.00e+00
2	1057	89.2	1206	28	T33161 Mutated OCIF, OCIF-C1	0.00e+00
3	1055	89.0	1206	27	T33165 Mutated OCIF, OCIF-C2	0.00e+00
4	1055	89.0	1206	27	T33164 Mutated OCIF, OCIF-C2	0.00e+00
5	1055	89.0	1206	28	T33162 Mutated OCIF, OCIF-C2	0.00e+00
6	1053	88.9	1206	27	T33163 Mutated OCIF, OCIF-CL	0.00e+00
7	1051	88.7	1200	27	T33172 Mutated OCIF, OCIF-CL	0.00e+00
8	1028	86.8	1182	27	T33178 Mutated OCIF, OCIF-CB	0.00e+00
9	1017	85.8	1173	28	T33175 Human tumour necrosis	0.00e+00
10	923	77.9	1083	27	T33166 Mutated OCIF, OCIF-DC	0.00e+00
11	907	76.5	1056	27	T33173 Mutated OCIF, OCIF-DC	0.00e+00
12	890	75.1	1080	27	T33167 Mutated OCIF, OCIF-DC	0.00e+00
13	806	68.0	966	27	T33179 Mutated OCIF, OCIF-CS	0.00e+00
14	765	64.6	1080	27	T33168 Mutated OCIF, OCIF-DC	0.00e+00
15	671	56.6	984	27	T33171 Mutated OCIF, OCIF-DC	0.00e+00

16	669	56.5	819	27	T33174	Mutated OCIF, OCIF-CD	0.00e+00
17	649	54.8	1080	27	T33169	Mutated OCIF, OCIF-DC	0.00e+00
18	453	38.2	981	27	T33170	Mutated OCIF, OCIF-DD	0.00e+00
19	444	37.5	594	27	T33175	Mutated OCIF, OCIF-CD	0.00e+00
20	413	34.9	564	27	T33180	Mutated OCIF, OCIF-CB	1.63e-288
21	391	33.0	10190	27	T33183	Fragment of human OCI	1.46e-271
22	287	23.9	432	27	T33176	Mutated OCIF, OCIF-CC	1.04e-188
23	263	22.5	321	27	T33177	Mutated OCIF, OCIF-CC	1.67e-176
24	265	22.4	438	28	T36689	Osteoclastogenesis in	5.58e-175
25	263	22.2	465	28	T36688	Osteoclastogenesis in	1.86e-173
26	246	20.8	255	27	T33181	Mutated OCIF, OCIF-CP	1.60e-160
27	88	7.4	1047	2	Q10572	Human Natriuretic Pep	4.90e-43
28	69	5.8	1047	2	Q10572	Human Natriuretic Pep	1.05e-29
29	45	3.8	91	9	Q51746	Oligonucleotide probe	1.16e-13
30	42	3.5	204	1	N81164	Base substituted E.co	9.24e-12
31	39	3.3	91	9	Q51746	Oligonucleotide probe	6.85e-10
32	39	3.3	204	1	N81164	Base substituted E.co	6.85e-10
33	35	3.0	114	12	Q70469	Generic DNA sequence	1.85e-07
34	35	3.0	114	12	Q70465	Generic DNA sequence	1.85e-07
35	34	2.9	114	12	Q70468	Generic DNA sequence	7.30e-07
36	34	2.9	114	12	Q70470	Generic DNA sequence	7.30e-07
37	33	2.8	114	12	Q70473	Generic DNA sequence	2.84e-06
38	33	2.8	114	12	Q70467	Generic DNA sequence	2.84e-06
39	32	2.7	114	12	Q70466	Generic DNA sequence	1.09e-05
40	32	2.7	114	12	Q70468	Generic DNA sequence	1.09e-05
41	32	2.7	114	12	Q70469	Generic DNA sequence	1.09e-05
42	32	2.7	114	12	Q70467	Generic DNA sequence	1.09e-05
43	32	2.7	114	12	Q70465	Generic DNA sequence	1.09e-05
44	30	2.5	114	12	Q70472	Generic DNA sequence	1.54e-04
45	30	2.5	114	12	Q70470	Generic DNA sequence	1.54e-04

## ALIGNMENTS

## RESULT 1

ID T36685 standard; DNA; 1206 BP.  
 AC T36685;  
 DT 22-APR-1997 (first entry)  
 DE Osteoclastogenesis inhibitory factor coding sequence.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT sig\_peptide 1..63  
 FT /\*tag= a  
 FT mat\_peptide 54..1203  
 FT /\*tag= b  
 FT /label= Claim 6  
 PN WO9626217-A1.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; J00374.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320-40.  
 DR P-PSDB: R99924-25.  
 PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 8; Page 66-67; 183pp; Japanese.  
 CC This sequence encodes the full length osteoclastogenesis inhibitory  
 CC factor (OCIF) of the invention. The OCIF has a molecular weight by  
 CC SDS-PAGE of 60 kD under reducing conditions and 120 kD under non-  
 CC reducing conditions. The protein is adsorbed onto cation-exchangers  
 CC or heparin and its activity is lowered after 10 mins at 70 deg.C or  
 CC 30 mins at 56 deg.C, and is lost after 10 mins at 90 deg.C. OCIF is  
 CC useful in the control of bone resorption and therefore in the  
 CC treatment and prevention of disorders of bone resorption, e.g.  
 CC osteoporosis.  
 SQ Sequence 1206 BP; 388 A; 284 C; 269 G; 265 T;  
 Query Match 89.4%; Score 1059; DB 28; Length 1206;





















Db	461	gtggaatagatgtttacccgtgtgtgaggagcattcttcagggttgctgttccatacaagtt	520
Qy	563	GTGGAAATAGATGTTTACCTGTGTGTGAGGAGCATCTTCTCAGGTTGTGCTGTCTACAAAGT	622
Db	521	ttacgcctaaactggcttagtcttggttagacaatttgctgtgacccaagaataaacgcag	580
Qy	623	TTAGCGCTTAACCTGGCTTAGTGTCTTGGTAGACAAATTGGCTGGCACCACCAAGTTAAACGCAG	682
Db	581	agagtgttagagagatataaacggccaacacagctcacagaacagactttccagctgtctga	640
Qy	683	AGAGTGTAGAGAGATATAACGGCAACACAGACTCACAAAGACAGACTTTCAGCTGCTGA	742
Db	641	agttatgaaacatccaacaagaaccaagatatagtcaagaagatcatccaagaatttg	700
Qy	743	AGTTATGGAAACATCAAAACAAGACCAGATATAGTCAAGAAGATCATCCCAAGATATTG	802
	701	acctctgtgaaaaacagcgtgcagcgccacttggacatgtaacctcaaccttcagcagc	760
	803	ACCTCTGTGAAAAACAGCGTGCAGCGGCACATTGGACATGCTAACCTCACCTTCGAGCAGC	862
Db	761	ttctgacttgatggaaagcttaccgggaaaaaagctggagacagaagacattgaaaaaa	820
Qy	863	TTCTGACTGTGATGGAAAGCTTTACCGGGAAAGAAAGTGGGGCAGCAAGACATTGAAAAA	922
Db	821	caataaaggcatgcaaacccagtgaccagatcctgaagctgtctagttgttggggaataa	880
Qy	923	CAATTAAGGGCATGCAAAACCCAGTGACAGATCTCTGAAGCTGCTCAGTTTGTGGCGAATAA	982
Db	881	aaaatgcgcaccaagacaccttgaaggcgctaatagcacgcactaaagcactcaaaagcgt	940
Qy	983	AAAAATGGCGAACCAAGACACCTTGAAGGGGCTTAATGCACGCATTAAGCAGCTCAAAAGACGT	1042
Db	941	accactttcccaaaactgtcactcagagtctaaagaagaccactcaggttctcttcacagct	1000
Qy	1043	ACCACCTTTCCCAAAACTGTTCACCTCAGAGTCTTAAGAAGACCATCAGGTTCTCTTCACAGCT	1102
Db	1001	tcacaatgtcaaatgtatcagaagtatttttagaataatgataggtaccaggttccaat	1060
Qy	1103	TCACAATGTACAATTTGATPCAGAAGTTATTTTTAGAAATGATAGGTAAACGAGTCCAAT	1162
Db	1061	cagtaaaaaataagctgttataa	1083
Qy	1163	CAGTAAAAAATAAGCTGCTTATAA	1185

Key	Location/Qualifiers
T33173 standard; DNA; 1056 BP.	
T33173;	
T33173;	
22-APR-1997 (first entry)	
DE Mutated OCIF, OCIF-CC, coding sequence.	
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;	
KW osteoporosis; ss.	
OS Synthetic.	
FT Key	
FT sig peptide	1..63
FT /*tag= a	
FT mat.peptide	64..1053
FT /*tag= b	
FT /product= OCIF-CC	
PN WO9626217-Al.	
PN 29-AUG-1996.	
PF 20-FEB-1996; J00374.	
PR 20-FEB-1995; JP-054977.	
PR 21-JUL-1995; JP-207508.	
PA (SNOW) SNOW BRAND MILK PROD CO LTD.	
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;	
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;	
WPI: 96-402320/40.	
P-PSDB: R99943.	
DR DNA encoding osteoclastogenesis inhibitory factor protein - useful	
PT for bone resorption control, esp. treatment of osteoporosis	
PS Claim 63; Page 144-145; 183pp; Japanese.	

```

|||||
QY 820 GTCCAGGCGCACATTGGACATGCTAACCTCACCTTCGAGCAGCTTCGTAGCTTGATGAA 879
Db 901 agttaccgggaaagaaagtgaggagcagagacattgaaaaaacaataaaagcagtcgaaa 960
QY 880 AGCTTACCGGGAAGAAAGTGGGAGCAGAGACATTGAAAAAACAAATAAAGGCATGCAAA 939
Db 961 cccagtgaccagatcctgaagctgcctcagcttctgagcgaataaaaaatggcgaccaagac 1020
QY 940 CCAGTGGACAGATCCTGAAGCTGCTCAGTTTGGCGAATAAAAAATGGCGACCAAGAC 999
Db 1021 acctgaaggggcctaagtcacgcactaaagcact 1054
QY 1000 ACCTGAAGGGCTAATGCACGCACCTAAAGCCT 1033

RESULT 12
ID T33167 standard; DNA; 1080 BP.
AC T33167;
DT 22-APR-1997 (first entry)
DE Mutated OCIF, OCIF-DCR2, coding sequence.
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;
OS Osteoporosis; ss.
FH Key Location/Qualifiers
FT sig_peptide 1..63
FT /*tag= a
FT mat_peptide 64..1077
FT /*tag= b
FT /product= OCIF-DCR2
PN WO9626217-A1.
PD 29-AUG-1996.
PF 20-FEB-1996; J00374.
PR 20-FEB-1995; JP-054977.
PR 21-JUL-1995; JP-207508.
PA (SNOW ) SNOW BRAND MILK PROD CO LTD.
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;
DR WPI: 96-402320/40.
DR P-PSDB: R99937.
PT DNA encoding osteoclastogenesis inhibitory factor protein - useful
  for bone resorption control, esp. treatment of osteoporosis
PS Claim 45; Page 138-139; 183pp; Japanese.
CC This sequence encodes a mutated version of the full length
  osteoclastogenesis inhibitory factor (OCIF) of the invention. This
  sequence encodes OCIF-DCR2 in which amino acids 43-84 of the mature
  protein have been deleted. The OCIF of the invention has a molecular
  weight by SDS-PAGE of 60 kD under reducing conditions and 120 kD under
  non-reducing conditions. The protein is adsorbed onto cation-exchangers
  or heparin and its activity is lowered after 10 mins at 70 deg.C or 30
  mins at 56 deg.C, and is lost after 10 mins at 90 deg.C. OCIF is useful
  in the control of bone resorption and therefore in the treatment and
  prevention of disorders of bone resorption, e.g. osteoporosis.
SQ Sequence 1080 BP; 357 A; 243 C; 236 G; 244 T;

Query Match 75.1%; Score 890; DB 27; Length 1080;
Best Local Similarity 99.9%; Pred.No. 0.00e+00;
Matches 891; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 189 cgaatgaaggaggcgctacattgagatagattctgttgaaacataggagctgcc 248
QY 294 CGAATGAAGGAGGCGCTACCTTGAGATAGAGTTCTGCTGAACATAGGAGCTGCC 353
Db 249 tcttgatttgagtgagtgcaagctggaacccagagcgaataacatagttgcaaaagatg 308
QY 354 TCCTGGATTTGGAGTGGTGGCAAGCTGGAAAGCCCGAGAGCAATACAGTTTGAAGAAGATG 413
Db 309 tccagatgggtcttctcaaatgagcgtcatctaagcacctgtagaacaacacaaa 368
QY 414 TCCAGATGGGTCTTCTCAANTAGAGCGTCATCTAAGCACCTGTAGAAAAACACACAAA 473
Db 369 ttgcagtgctttgttctctgcttaactcagaagaaatgcaacacgacgaacatatg 428
|||||

```

```

QY 474 TTGCAGTGTCTTTGGTCTCTCTGCTAACTCAGAAAGAAATGCAACACGACACACATATG 533
Db 429 ttccggaaaacagtgaaatcaactcaaaaatgtggaatagattgttaacctgtgtgagagagc 488
QY 534 TTCGGGAAACAGTGAATCAACTCAAAATGTGGAATAGATGTTACCTCTGTGAGGAGGC 593
Db 489 attcttcaggttttgctgttctcacaagtttacgcctaaactgagcttagtgccttctgtaga 548
QY 594 ATTCTTCAGGTTTGCTGTCTCTACAAAGTTTACCCCTAACTGGCTTAGTGTCTTGGTAGA 653
Db 549 caatttcctggcccaaaagtaaacgcagagagtgtagagagataaaaaacggcgaacacag 608
QY 654 CAATTTCCTTGGCACCACAAAGTAAACGCAGAGAGTGTAGAGAGATAAAACGGCAACACAG 713
Db 609 ctcaagaagacagactttccagctgctgaagttatagaaacatcaaaaacaaagaccaaga 668
QY 714 CTCACAAGAACAGACTTTCAGCTGCTGAAGTTATGAAACATCAAAACAAAGACCAAGA 773
Db 669 tatagtcaagaagatcatccaagatatgtacctctgtgaaaaacagcgtgcagcggcacat 728
QY 774 TATAGTCAAGAAGATCATCAAGATATTGACCTCTGTGAAAAACAGCGTGCAGCGGCACAT 833
Db 729 tggacatgctaactcactccttcagcagcgttcgtagctgtagtgaagctttaccgggaaa 788
QY 834 TGGACATGCTAACCTCACCTTCGAGCAGCTTCGTAGCTTCATGGAAGCTTACC GGAAA 893
Db 789 gaaagtgggagcagaagacattgaaaaaacaataaaggcagtgcaaacccagtgaccagat 848
QY 894 GAAAGTGGGAGCAGAGACATTGAAAAAACAAATAAAGGCATGCAAAACCCAGTACCAGAT 953
Db 849 cctgaagctgctcagttgtggtggaataaaaaatggcgaccaagacacacctgaaggccct 908
QY 954 CCTGAAGCTGCTCAGTTTGTGGCGAATAAAAAATGGGACCAAGACACACTTGAAGGCGCT 1013
Db 909 aatgcacgcactaaagcactcaaaagcagctaccactttcccaaacctgtaactcagagtct 968
QY 1014 AATGCACGCCTAAAGCACCTCAAGAGCTACCACCTTTCCCAAACTGTCTACTCAGAGTCT 1073
Db 969 aaagaagacacatcaggttcttcacagcttcacaaatgtacaaatgttatcacaaagtatt 1028
QY 1074 AAGAAGAACCATCAGTTCCTTCACAGCTTCACAATGTACAAATGTATCAGAAGTTATT 1133
Db 1029 tttagaatgataggttaaccaggtcccaatcagtaaaaaataagctgcttataa 1080
QY 1134 TTAGAATAAGTAGTAAACCAGGTCCCAATCAGTAAAAAATAAGTCTGCTTATAA 1185

RESULT 13
ID T33179 standard; DNA; 966 BP.
AC T33179;
DT 22-APR-1997 (first entry)
DE Mutated OCIF, OCIF-CSph, coding sequence.
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;
OS Osteoporosis; ss.
FH Key Location/Qualifiers
FT sig_peptide 1..63
FT /*tag= a
FT mat_peptide 64..963
FT /*tag= b
FT /product= OCIF-CSph
PN WO9626217-A1.
PD 29-AUG-1996.
PF 20-FEB-1996; J00374.
PR 20-FEB-1995; JP-054977.
PR 21-JUL-1995; JP-207508.
PA (SNOW ) SNOW BRAND MILK PROD CO LTD.
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;
DR WPI: 96-402320/40.
DR P-PSDB: R99949.
PT DNA encoding osteoclastogenesis inhibitory factor protein - useful
  for bone resorption control, esp. treatment of osteoporosis

```

PS Claim 81; Page 149; 183pp; Japanese.  
 CC This sequence encodes a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC sequence encodes OCIF-C39h in which amino acids 298-380 of the mature  
 CC protein have been deleted and replaced by Ser-Leu-Asp. These amino  
 CC acid changes have been caused by the introduction of a restriction  
 CC site. The OCIF of the invention has a molecular weight by SDS-PAGE of  
 CC 60 kD under reducing conditions and 120 kD under non-reducing  
 CC conditions. The protein is adsorbed onto cation-exchangers or heparin  
 CC and its activity is lowered after 10 mins at 70 deg.C or 30 mins at 56  
 CC deg.C, and is lost after 10 mins at 90 deg.C. OCIF is useful in the  
 CC control of bone resorption and therefore in the treatment and  
 CC prevention of disorders of bone resorption, e.g. osteoporosis.  
 SQ Sequence 966 BP; 301 A; 228 C; 226 G; 211 T;

Query Match 68.0%; Score 806; DB 27; Length 966;  
 Best Local Similarity 97.8%; Pred. No. 0.00e+00;  
 Matches 932; Conservative 0; Mismatches 0; Indels 21; Gaps 1;  
 1 atgaacaacttgctgtgcgcgcgtctgtttctggacatctccattaaagtgaccacc 60  
 1 ATGAACAACCTTGTGTGTCGCGCTCGTGTTCCTGGACATCTCCATTAAAGTGGACCA 60  
 61 caggaaactttctctccaaagtaactcttattatgacgaaagaaacctctcaagctgtg 120  
 61 CAGGAAACCTTTCCTCCAAAGTACCTTCTATTATGACGAAAGAACCTCTCATCAGCTGTG 120  
 121 tgtgacaaatgtctctctggtacctaccctaaacaaacactgtacagcaagtggaagacc 180  
 121 TGTGACAAATGTCTCTGCTGCTACCTACCTAAACAAACACTGTACAGCAAGTGGAGACC 180  
 181 gtgtgccccctgcccctgaccactactacacagacagctggcacaccagtgagagtg 240  
 181 GTGTGCCCCCTTGCCCTGACCACTACTACACAGACAGCTGGCACACCAGTGAAGTGT 240  
 241 ctatactgcagcccgctgtgcaagagctgcagtcagtcagtcagtcagtcagtcagtc 300  
 241 CTATCTGTCAGCCCGCTGTGCAAGGAG-----TGCAATGCGCACC 279  
 301 cacaacgcgtgtggaatgcgaaggaaggcgtaccttgagatagatgttctgttgaaa 360  
 301 CACAACGCGTGTGGAATGCAAGGAGGCGCTACCTTGAGATAGATGCTGTGTTGAAA 339  
 361 catagagctgcctctctgattgttgagtggtgcgaagctggaaccccgagcgaataca 420  
 361 CATAGAGCTGCCCTCTCTGATTTGGAGTGTGCAAGTGGAAACCCCGAGCGGAATACA 399  
 421 gttgcaaaagatgtccagatgggttctctcaaatgagacgtcatctaaagcaccctgt 480  
 421 GTTGTGCAAAAGATGTCCAGATGGGTCTCTCAAAATGAGACGTCACTAAAGCACCCCTGT 459  
 481 agaaacacacaaattgcagtgctcttgggtctcctgcttaactcagaaggaagaaatgcaaca 540  
 481 AGAAAAACACAAATGTCAGTGTCTTGGTCTCTGCTTAACCTCAGAAAGGAATGCAACA 519  
 541 cagcaacaatatgttcggaaacagtaagtaactcaactcaaaatgtggaatgagtgtaacc 600  
 541 CAGCAACAATATGTTCCGGAAACAGTGAATCAACTCAAAATGTGGAATGATGTTTACC 579  
 601 ctgtgtgaggagcattcttcagggttctctcaaatgagacgtttacgccttaactggcct 660  
 601 CTGTGTGAGGAGCATCTTTCAGGTTTGTCTGTCTCTCAAAAGTTTACGCCCTTAACCTGCTT 639  
 661 agtgtttgtgagcaattgtcctggccccaagtaaacgcagagagtgtagagagata 720  
 661 AGTGTCTTGTGTAGCAATTTGCTGCTGGCCCAAGTAAACGCGAGAGTGTAGAGAGGATA 699  
 721 aaacggcaacagctcacaagaacagactttccagctgctgaagttaagaaacatcaa 780  
 721 AAACGGCAACAGCTCACAAGAACAGACTTTCCAGCTGCTGGAAGTTAAGGAACATCAA 759  
 781 aacaagaccaagatatagtcgaagaagatcatcccaagatatattgacctctgtgaacaacagc 840

QY 760 AACAAAGACCAAGATATAGTCAAGAAGATCATCCAAAGATATTGACCTCTGTGAAACACG 819  
 Db 841 gtccagcggcacattggacatgcttaacctcaactctgacagcagcttgtagtgtaa 900  
 QY 820 GTCCAGCGGCACATTGGACATGCTAACCTCACCTTGGACAGCTTGTGATGGAA 879  
 Db 901 agcttaccgggaaagaaagtggagcagagacattgaaaaaacaataaagc 953  
 QY 880 AGCTTACCGGGAAGAAAGTGGGAGCAGAGACATTTGAAAAACAATAAAGGC 932  
 RESULT 14  
 ID T33168 standard; DNA; 1080 BP.  
 AC T33168;  
 DT 22-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-DCR3, coding sequence.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT sig\_peptide 1..63  
 FT /\*tag= a  
 FT mat\_peptide 64..1077  
 FT /\*tag= b  
 FT /product= OCIF-DCR3  
 PN WO9626217-A1.  
 PD 29-AUG-1996.  
 PF 20-FEB-1996; J00374.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR P-PSDB; R9938.  
 PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 48; Page 139-140; 183pp; Japanese.  
 CC This sequence encodes a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC sequence encodes OCIF-DCR3 in which amino acids 85-122 of the mature  
 CC protein have been deleted. The OCIF of the invention has a molecular  
 CC weight by SDS-PAGE of 60 kD under reducing conditions and 120 kD under  
 CC non-reducing conditions. The protein is adsorbed onto cation-exchangers  
 CC or heparin and its activity is lowered after 10 mins at 70 deg.C or 30  
 CC mins at 56 deg.C, and is lost after 10 mins at 90 deg.C. OCIF is useful  
 CC in the control of bone resorption and therefore in the treatment and  
 CC prevention of disorders of bone resorption, e.g. osteoporosis.  
 SQ Sequence 1080 BP; 351 A; 259 C; 233 G; 237 T;  
 Query Match 64.6%; Score 765; DB 27; Length 1080;  
 Best Local Similarity 100.0%; Pred. No. 0.00e+00;  
 Matches 765; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 316 agaagtcacagatgggtctctcctcaaatgagacgtcatctaaagcaccctgtagaacaac 375  
 409 AGATGTCCAGATGGGTCTCTCAAAATGAGACGTCTATTAAGCACCCCTGTAGAAAACAC 468  
 376 acaaaatgcagtgcttctgtctcctgctcaactcagaaggaatgcaacacacacaac 435  
 469 ACAAAATGCAGTGCTCTTGGTCTCTGCTTAACCTCAGAAAGGAATGCAACACACCAAC 528  
 436 atatgttccggaaacagtgaaatcaactcaaaatgtggaatagattaccctgtgtgag 495  
 529 ATATGTTCGGAAACAGTGAATCAACTCAAAATGTGGAATAGATGTTACCTGTGTGAG 588  
 496 gaggcattctcaggttctctcctcaaaagtttacgccttaactggttagttcttg 555  
 589 GAGGCATTCTCAGGTTTGTCTTCTTCTTACAAAGTTTACGCCCTTAACCTGTAGTCTGTG 648  
 556 gtacacaatttccctggccaccaagtaaacgcagagagtgtagagagataaaacggcaa 615  
 649 GTAGACAATTTGCTGGCCACCAAGTAAGTAACGCGAGAGTGTAGAGAGGATAAACGGCAA 708

Db 616 cacagctcaaaagaacagactctccagctgctgaagtattggaacatcaaaacaaagac 675  
 QY 709 CACAGCTCACAGAACAGACTTCCAGCTGCTGAAGTTATGGAACATCAAAACAAAGAC 768  
 Db 676 caagatatagtcagaagatcatccaaagatatgaactctgtgaaacagcgtgcagcg 735  
 QY 769 CAAGATATAGTCAAGAAGATCATCCAAAGATATTGACCTCTGTGAAACAGCGTGCAGCG 828  
 Db 736 cacattgagatgtaactcacttcagcagcttcgttagcttgatgaaagcttacgg 795  
 QY 829 CACATTGGACATGTAACCTACCTTCAGCAGCTTCGTAGCTTGATGGAACCTTACCG 888  
 Db 796 ggaagaagtgaggagcagaagacattgaaacaaataaagcagtcgaaacccagtgac 855  
 QY 889 GGAAGAAGTGGGAGCAGAGACATTGAAAAACAATAAAGGCATGCAAAACCCAGTGAC 948  
 Db 856 cagatctgaagctgctcagtttggcgaaataaaaaatggcgacacacacttgaag 915  
 QY 949 CAGATCTGAAGCTGCTCAGTTTGTGGGAATAAAAAATGGCGACCAAGACACTTGAAG 1008  
 Db 916 ggcctaagtcagcactaaagcactcaaaagcgtaccactttcccaaaactgtcactcag 975  
 QY 1009 GGCCTAATGACGACATAAAGCATTCAAGAGCTTACCACITTCCTCAAACTGTCACTCAG 1068  
 Db 976 agtcaagaagaccatcaggttctctccacagcttcacaaatgtacaaattgtatcagaag 1035  
 QY 1069 AGTCTAAAGAAGACCATCAGGTTCCTTCCACAGCTTCAAAATGTACAAATGTATCAGAAG 1128  
 Db 1036 ttatttttagaattagatgtaaacaggtccaatcagtaaaaaata 1080  
 QY 1129 TTATTTTGTAGAAATGATAGGTAAACAGGTCCCAATCAGTAAATAAATA 1173

## RESULT 15

ID T33171 standard; DNA; 984 BP.  
 AC T33171:  
 DT 22-APR-1997 (first entry)  
 DE Mutated OCIF, OCIF-DDD2, coding sequence.  
 KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
 KW osteoporosis; ss.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT sig\_peptide 1..63  
 FT /\*tag= a 64..981  
 FT mat\_peptide 64..981  
 FT /\*tag= b  
 FT /product= OCIF-DDD2  
 PN W09626217-A1.  
 DT 29-AUG-1996.  
 PR 20-FEB-1996; J00374.  
 PR 20-FEB-1995; JP-054977.  
 PR 21-JUL-1995; JP-207508.  
 PA (SNOW ) SNOW BRAND MILK PROD CO LTD.  
 PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
 PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
 DR WPI: 96-402320/40.  
 DR P-PSDB; R99941.  
 PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
 PT for bone resorption control, esp. treatment of osteoporosis  
 PS Claim 57: Page 142-143; 183pp; Japanese.  
 CC This sequence encodes a mutated version of the full length  
 CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
 CC sequence encodes OCIF-DDD2 in which amino acids 233-326 of the mature  
 CC protein have been deleted. The OCIF of the invention has a molecular  
 CC weight by SDS-PAGE of 60 kD under reducing conditions and 120 kD under  
 CC non-reducing conditions. The protein is adsorbed onto cation-exchangers  
 CC or heparin and its activity is lowered after 10 mins at 70 deg.C or 30  
 CC mins at 56 deg.C, and is lost after 10 mins at 90 deg.C. OCIF is useful  
 CC in the control of bone resorption and therefore in the treatment and  
 CC prevention of disorders of bone resorption, e.g. osteoporosis.  
 CC Sequence 984 BP; 313 A; 236 C; 212 G; 223 T;

Query Match 56.6%; Score 671; DB 27; Length 984;

Best Local Similarity 97.4%; Pred. No. 0.00e+00;

Matches 797; Conservative 0; Mismatches 0; Indels 21; Gaps 1;

Db 1 atgaacaaactgtgtgctgcgcgtcgtgttcttgacatctccatttaagtggaccacc 60  
 QY 1 ATGAACAACACTTGTGTGCTGCGCGCTGCTGTTCTTGGACATCTCCATTAAAGTGACCCACC 60  
 Db 61 caggaaacgttctctccaaagtacaccttcattatgacgaagaacacctctcatcagctgttg 120  
 QY 61 CAGGAACAGTTTCTCTCCAAAGTACCTTTCATTATGACGAAGAAGCTCTCTATCAGCTGTG 120  
 Db 121 tctgaacaaatgtctctgtgtacacctcaaaacaaacactgtacagcaaaagtgaagacc 180  
 QY 121 TGTGACAAATGTCCTTGGTACCTTACCTTAAACAACACTGTACAGCAAGTGAAGAC 180  
 Db 181 gttgtgcgccttgccttgaccactactacacagacagcagctggcacaccagtgagagtg 240  
 QY 181 GTGTGCGCCCTTGCCTTGACCCTACTTACAGACAGCTGGCACACCCAGTGACGAGTGT 240  
 Db 241 ctatactgcagcccgctgtgcaaggagctgcagtcagtcgaagcaggagtgcaatcgacc 300  
 QY 241 CTATACTGCAGCCCGCTGTGCAAGGAG-----TGCATTCGCRACC 279  
 Db 301 caaacccgctgtgcgaatgcaaggagggcgctaccttgagatagagttcttctgttgaaa 360  
 QY 280 CACAACCGCTGTGGAATGCAAGGAAGGGCTACTTGTGATAGACTTCTGCTTGAAA 339  
 Db 361 cataggagctgcctctccttgattgtgagtggtgcgaagtggaaaccccgagcgaaataca 420  
 QY 340 CATAGGAGCTGCCCTCCTTGGATTGAGTTGGTGTGCAAGCTGGAACCCAGAGCGAATAACA 399  
 Db 421 gtttgcaaaagatgtccagatggtgtctctcaaatgagcgtcatctaaagcaccctgt 480  
 QY 400 GTTTGCAAAAGATGTCCAGATGGGTCTTCTCAAAATGAGACGTCACTAAACACCCCTGT 459  
 Db 481 agaaacacacaaaattgcagtgctcttggctcctgtctaaactcagaagaagaatgcaaca 540  
 QY 460 AGAAACACACAAATGTCAGTGTCTTTGGTCTCCTGCTAACTCAGAAAGAAATGCAACA 519  
 Db 541 caccgacaacatagtgtcccgaaacagtgtaaatcaactcaaaatgtggaatagattacc 600  
 QY 520 CACGACAACATATGTTCCGGAACACAGTGAATCAACTCAAAAATGTGGAATAGATTACC 579  
 Db 601 ctgtgtgagaggcattcttcaggtttgtctcctcaaaagttaacgcttaactggctt 660  
 QY 580 CTGTGTGAGGAGGCAITCTTCAGTTTGTGTCTCTACAAAGTTTACGCTTAACGTGCTT 639  
 Db 661 agtgtctgttagacaatttgcctggcaccacaagttaaacgcagagagtgtagaggata 720  
 QY 640 AGTGTCTTGTGTAGACAATTTGCTGCGCACCAAGTAACGACAGAGAGTGTAGAGAGGATA 699  
 Db 721 aaacggcaacacagctcacaagaacacacttccagctcgaagttaaggaaacatcaa 780  
 QY 700 AAACGGCAACACAGCTCACAAGAACACAGCTTCCACGCTGCTGAAGTTATGGAACATCAA 759  
 Db 781 aacaagaccacaagatatagtccaagaagatcatccaaga 818  
 QY 760 AACAAAGACCCAAAGATATAGTCAAGAGATCATCTCAAGA 797

Search completed: Tue Dec 2 15:25:57 1997

Job time : 346 secs.







SOURCE  
ORGANISM

human.  
Homo sapiens  
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;  
Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 602)  
Hallier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and  
Wilson, R.  
The WashU-Merck EST Project  
Unpublished (1995)

REFERENCE  
AUTHORS

Contact: Wilson RK  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 1203 Std Error: 0.00  
Seq primer: -28M13 rev2 from Amersham  
High quality sequence stop: 435.  
Location/Qualifiers  
1..602  
/organism="Homo sapiens"  
/note="Vector: pT73D (Pharmacia) with a modified  
polylinker V-NTPE; phagemid: Site\_1; Not I; Site\_2: Eco  
RI; TGTTACCAATCTGAAGGCGGCCGCAATTTTTTTTTTTTTT  
3'], double-stranded cDNA was size selected, ligated to  
Eco RI adapters (Pharmacia), digested with Not I and  
cloned into the Not I and Eco RI sites of a modified pT713  
vector (Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Bento  
Soares and M Fatima Bonaldo."  
/clone="325973"  
/clone\_lib="Soares senescent fibroblasts NBHSF"  
/tissue\_type="senescent fibroblast"  
/lab\_host="DH10B (ampicillin resistant)"  
<1..>602

FEATURES  
source

209 a 128 c 136 g 126 t 3 others  
Query Match 47.5%; Score 563; DB 195; Length 602;  
Best Local Similarity 98.7%; Pred. No. 0.00e+00;  
Matches 588; Conservative 0; Mismatches 4; Indels 4; Gaps 4;  
Db 2 tctaaagcaccctgtagaacacacacaaatttcgcagtgcttggctcctgtaactca 61  
|||||  
QY 445 TCTAAGCACCTGTAGAAAACACACAAATT-GCAGTGTCTTTGGTCTCTCTGCTAACTCA 503  
|||||  
Db 62 gaaggaatgcaacacacagcaacatatgttccggaacagtgtaacacactcaaaatg 121  
|||||  
QY 504 GAAAGGAATGCAACACAGCAACATATGTTCGGAAACAGTGAATCAACTCAAAATG 563  
|||||  
Db 122 tggatagatgtaccctgtgtgaggagcaattttcagggtttgtgttctcacaaggt 181  
|||||  
QY 564 TGAATAGATGTATACCTGTGTGAGGAGGCAATCTTCAGGTTTGTGTCTCTACAAAGTT 623  
|||||  
Db 182 tacgcttaactggcttagtctgttgtagacaatttgcctggcaccacaaagtaaacgaga 241  
|||||  
QY 624 TACGCTTAATGGCTTAGTGTCTTGGTAGCAATTTGCTGGCACCACAAAGTAAACGAGA 683  
|||||  
Db 242 gaggttagagaggtataaacggcaacacagcttccaaagaacagactttccagctgctgaa 301  
|||||  
QY 684 GAGTGTAGAGAGGATAAAGCGCAACACAGCTCAACAGACAGACTTTCAGCTGCTGAA 743  
|||||  
Db 302 gttatggaacatcaaaacaaagacacagatatagtcagaagatcatccaaagatatga 361  
|||||  
QY 744 GTTATGGAACATCAAAACAAAGACCAAGATATAGTCAAGAAGATCATCCAAGATATGA 803  
|||||

BASE COUNT  
ORIGIN

209 a 128 c 136 g 126 t 3 others  
Query Match 47.5%; Score 563; DB 195; Length 602;  
Best Local Similarity 98.7%; Pred. No. 0.00e+00;  
Matches 588; Conservative 0; Mismatches 4; Indels 4; Gaps 4;  
Db 2 tctaaagcaccctgtagaacacacacaaatttcgcagtgcttggctcctgtaactca 61  
|||||  
QY 445 TCTAAGCACCTGTAGAAAACACACAAATT-GCAGTGTCTTTGGTCTCTCTGCTAACTCA 503  
|||||  
Db 62 gaaggaatgcaacacacagcaacatatgttccggaacagtgtaacacactcaaaatg 121  
|||||  
QY 504 GAAAGGAATGCAACACAGCAACATATGTTCGGAAACAGTGAATCAACTCAAAATG 563  
|||||  
Db 122 tggatagatgtaccctgtgtgaggagcaattttcagggtttgtgttctcacaaggt 181  
|||||  
QY 564 TGAATAGATGTATACCTGTGTGAGGAGGCAATCTTCAGGTTTGTGTCTCTACAAAGTT 623  
|||||  
Db 182 tacgcttaactggcttagtctgttgtagacaatttgcctggcaccacaaagtaaacgaga 241  
|||||  
QY 624 TACGCTTAATGGCTTAGTGTCTTGGTAGCAATTTGCTGGCACCACAAAGTAAACGAGA 683  
|||||  
Db 242 gaggttagagaggtataaacggcaacacagcttccaaagaacagactttccagctgctgaa 301  
|||||  
QY 684 GAGTGTAGAGAGGATAAAGCGCAACACAGCTCAACAGACAGACTTTCAGCTGCTGAA 743  
|||||  
Db 302 gttatggaacatcaaaacaaagacacagatatagtcagaagatcatccaaagatatga 361  
|||||  
QY 744 GTTATGGAACATCAAAACAAAGACCAAGATATAGTCAAGAAGATCATCCAAGATATGA 803  
|||||

Db 362 cctctgtgaaaaacagcgtgcagcgccacattggagacatgctaaacctcacccttcagcagc 421  
|||||  
QY 804 CCTCTGTGAAAAACAGCGTGCAGCGGCACATGG-ACATGCTAACCTCACCTTCGAGCAGC 862  
|||||  
Db 422 ttctgtagcttgatggaagccttaccgggggaagaaagtgaggagcagaagacattgaaaaa 481  
|||||  
QY 863 TT-CGTAGCTTGATGGAAGACTTACCGGGGAAAGAAAGTGGGACGACAGACATTTGAAAA 921  
|||||  
Db 482 acaataaagcagtcgcaaacccagtcgaccgatnctcagagctcagctgtctgttggcgcaata 541  
|||||  
QY 922 ACAATAAAGGCATGCAAAACCCAGTCAGCAGATCCTGAAGCTGCTCAGTTTGTGGCAATA 981  
|||||  
Db 542 aaaaatggcgaccacaaganaccttgaagggncataatgcacgcaacttaaacactcaa 597  
|||||  
QY 982 AAAAATGGCGACCAAGACACCTTGAA-GGGCCTAATGACGACGACTAAAGCACTCAA 1036  
|||||

## RESULT

2 H88769 346 bp mRNA EST 22-NOV-1995  
LOCUS yw23q12.r1 Homo sapiens CDNA clone 253126 5'  
DEFINITION H88769  
ACCESSION H88769  
NID 91071029  
KEYWORDS EST.  
SOURCE human clone=253126 primer=M13RP1 library=Morton Fetal Cochlea  
vector=pBluescript SK- host=SOLR cells (kanamycin resistant)  
XhoI-EcoRI Rsite2-XhoI The cDNA was oligo (dT) primed with an  
XhoI restriction enzyme recognition site and an 18 base poly dT  
sequence. For the 5' end, the synthesized cDNA termini were  
treated with 14 DNA polymerase and EcoRI adaptors were ligated to  
the blunt ends. adaptor linker: GAAATTCGGCAGGAG.

## ORGANISM

Homo sapiens  
Eukaryotae; Metazoa; Emnetazoa; Bilateria; Coelomata;  
Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;  
Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;  
Eutheria; Archonta; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 346)  
Hallier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevaskis, E.,  
Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.  
WashU-Merck EST Project  
Unpublished (1995)

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT









Qy 1117 TTGTATCAGAAAGTTATTTTGTAGAAATGA 1144

RESULT 11  
 LOCUS HSDH23G03 387 bp RNA EST 03-FEB-1994  
 DEFINITION H. sapiens partial cDNA sequence; clone H23G03; single read.  
 ACCESSION 221693  
 NID g38548  
 KEYWORDS partial cDNA sequence; transcribed sequence fragment.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE  
 1 (bases 1 to 387)  
 Eukaryotes; Mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo. Genexpress.

TITLE Direct Submission

JOURNAL Submitted (20-FEB-1993) Genzentrum Muenchen, Laboratorium fuer molekulare Biologie, Am Klopferspitz 18a, 8033 Martinsried, Germany. E-mail: obermaier@ms.biochem.mpg.de

REFERENCE  
 2 (bases 1 to 387)  
 Genexpress.

TITLE The Genexpress cDNA program

JOURNAL Unpublished

COMMENT Cloning vector is pBluescript SK(+); Genexpress library reference is A. automatic.

FEATURES  
 source Location/Qualifiers  
 1..387  
 /organism="Homo sapiens"  
 /clone\_lib="Stratagene cDNA library Human heart, cat#936208"

BASE COUNT 106 a 94 c 78 g 109 t

ORIGIN

Query Match 1.9%; Score 22; DB 109; Length 387;  
 Best Local Similarity 86.7%; Pred. No. 2.66e-04;  
 Matches 26; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 254 cagtaccacaaacctgaacctgtcagttt 283  
 ||||| | ||||| ||||| ||||| |||||  
 Qy 942 CAGTGACGAGATCTGAGGTCGTCAGTTT 971

RESULT 12  
 LOCUS W76826 406 bp mRNA EST 20-JUN-1996  
 DEFINITION me73g11.r1 Soares mouse embryo NbMEL3.5 14.5 Mus musculus cDNA clone 401252 5'.  
 ACCESSION W76826  
 NID g1387900  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus

REFERENCE  
 1 (bases 1 to 406)  
 Eukaryotes; Mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae; Murinae; Mus.  
 AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE The WashU-HMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mousest@wustl.wustl.edu  
 This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:245020  
 Putative full length read  
 Seq primer: ETPrimer  
 High quality sequence stop: 349.  
 Location/Qualifiers  
 1..406  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAGTGGAGCGCGCGGAAATTTTTTTTTTTTTTT T 3'], on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru KO, Wayne State Univ., from 2 ]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."  
 /clone="401252"  
 /clone\_lib="Soares mouse embryo NbMEL3.5 14.5"  
 /sex="unknown"  
 /tissue\_type="embryo"  
 /dev\_stage="13.5-14.5dpc total fetus"  
 /lab\_host="DH10B"  
 <1..>406

BASE COUNT 111 a 105 c 93 g 97 t

ORIGIN

Query Match 1.9%; Score 22; DB 126; Length 406;  
 Best Local Similarity 72.0%; Pred. No. 2.66e-04;  
 Matches 36; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Db 355 agctgaccgaggaagaaataaaagagccagtcattgaaaaa 404  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Qy 880 AGCTTACCGGAGAAAGTGGGAGCAGACATTTGAAAAACATATA 929

RESULT 13  
 LOCUS T74804 410 bp mRNA EST 02-MAR-1995  
 DEFINITION yc60d04.r1 Homo sapiens cDNA clone 85063 5'.  
 ACCESSION T74804  
 NID 9691479  
 KEYWORDS EST.  
 SOURCE human clones=85063 library=Stratagene liver (#937224) vector=pBluescript SK host=SOLR cells (kanamycin resistant) primer=M13RPL Rsite1-EcoRI Rsite2=XhoI Cloned unidirectionally.  
 Primer: Oligo dT. Hepatotomy from normal 49 year old male caucasian. Average insert size: 1.1 kb; Uni-ZAP XR Vector; 5' adaptor sequence: 5'-GAATTCGCGACGAG-3'; 3' adaptor sequence: 5'-CTCGAGTTTTTTTTTTTTTTT-3'.

ORGANISM Homo sapiens

REFERENCE  
 1 (bases 1 to 410)  
 Eukaryotes; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Fultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

TITLE WashU-Merck EST Project

JOURNAL Unpublished (1995)

COMMENT Contact: Wilson RK  
 WashU-Merck EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.wustl.edu  
 High quality sequence stops: 212  
 Source: IMAGE Consortium, LLNL  
 This clone is available royalty-free through LLNL; contact the





\*\*\*\*\*

WQSEFH (TM)

\*\*\*\*\*

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

MPsrch\_nnn n.a. - n.a. database search, using Smith-Waterman algorithm  
Run on: Tue Dec 2 15:51:34 1997; MasPar time 325.28 Seconds  
Parallel output not generated. 1001.129 Million cell updates/sec

File: >US-08-915-004-8  
Description: (1-1185) from US08915004.seq  
Perfect Score: 1185  
N.A. Sequence: 1 ATGACACTTCTGCTGCTG.....TAAATAAGCTGCTTAA 1185  
Comp: TACTGTGTAACACACACAC.....ATTTTATTCACGAAATAT

Scoring table: TABLE default  
Gap 6  
Nmatch STD : Dbase 0; Query 0  
Searched: 359085 seqs, 137405154 bases x 2  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: EST-STS-THREE  
1:EST199 2:EST200 3:EST201 4:EST202 5:EST203 6:EST204  
7:EST205 8:EST206 9:EST207 10:EST208 11:EST209 12:EST210  
13:EST211 14:EST212 15:EST213 16:EST214 17:EST215  
18:EST216 19:EST217 20:EST218 21:EST219 22:EST220  
23:EST221 24:EST222 25:EST223 26:EST224 27:EST225  
28:EST226 29:EST227 30:EST228 31:EST229 32:EST230  
33:EST231 34:EST232 35:EST233 36:EST234 37:EST235 38:EST236  
39:EST237 40:EST238 41:EST239 42:EST240 43:EST241 44:EST242 45:EST243  
46:EST244 47:EST245 48:EST246 49:EST247 50:EST248 51:EST249 52:EST250 53:EST251  
54:EST252 55:EST253 56:EST254 57:EST255 58:EST256 59:EST257 60:EST258 61:EST259 62:EST260 63:EST261  
64:EST262 65:EST263 66:EST264 67:EST265 68:EST266 69:EST267 70:EST268 71:EST269 72:EST270 73:EST271 74:EST272 75:EST273 76:EST274 77:EST275 78:EST276 79:EST277 80:EST278 81:EST279 82:EST280 83:EST281 84:EST282 85:EST283 86:EST284 87:EST285 88:EST286 89:EST287 90:EST288 91:EST289 92:EST290 93:EST291 94:EST292 95:EST293 96:EST294 97:EST295 98:EST296 99:EST297 100:EST298 101:EST299 102:EST300 103:EST301 104:EST302 105:EST303 106:EST304 107:EST305 108:EST306 109:EST307 110:EST308 111:EST309 112:EST310 113:EST311 114:EST312 115:EST313 116:EST314 117:EST315 118:EST316 119:EST317 120:EST318 121:EST319 122:EST320 123:EST321 124:EST322 125:EST323 126:EST324 127:EST325 128:EST326 129:EST327 130:EST328 131:EST329 132:EST330 133:EST331 134:EST332 135:EST333 136:EST334 137:EST335 138:EST336 139:EST337 140:EST338 141:EST339 142:EST340 143:EST341 144:EST342 145:EST343 146:EST344 147:EST345 148:EST346 149:EST347 150:EST348 151:EST349 152:EST350 153:EST351 154:EST352 155:EST353 156:EST354 157:EST355 158:EST356 159:EST357 160:EST358 161:EST359 162:EST360 163:EST361 164:EST362 165:EST363 166:EST364 167:EST365 168:EST366 169:EST367 170:EST368 171:EST369 172:EST370 173:EST371 174:EST372 175:EST373 176:EST374 177:EST375 178:EST376 179:EST377 180:EST378 181:EST379 182:EST380 183:EST381 184:EST382 185:EST383 186:EST384 187:EST385 188:EST386 189:EST387 190:EST388 191:EST389 192:EST390 193:EST391 194:EST392 195:EST393 196:EST394 197:EST395 198:EST396 199:EST397 200:EST398 201:EST399 202:EST400 203:EST401 204:EST402 205:EST403 206:EST404 207:EST405 208:EST406 209:EST407 210:EST408 211:EST409 212:EST410 213:EST411 214:EST412 215:EST413 216:EST414 217:EST415 218:EST416 219:EST417 220:EST418 221:EST419 222:EST420 223:EST421 224:EST422 225:EST423 226:EST424 227:EST425 228:EST426 229:EST427 230:EST428 231:EST429 232:EST430 233:EST431 234:EST432 235:EST433 236:EST434 237:EST435 238:EST436 239:EST437 240:EST438 241:EST439 242:EST440 243:EST441 244:EST442 245:EST443 246:EST444 247:EST445 248:EST446 249:EST447 250:EST448 251:EST449 252:EST450 253:EST451 254:EST452 255:EST453 256:EST454 257:EST455 258:EST456 259:EST457 260:EST458 261:EST459 262:EST460 263:EST461 264:EST462 265:EST463 266:EST464 267:EST465 268:EST466 269:EST467 270:EST468 271:EST469 272:EST470 273:EST471 274:EST472 275:EST473 276:EST474 277:EST475 278:EST476 279:EST477 280:EST478 281:EST479 282:EST480 283:EST481 284:EST482 285:EST483 286:EST484 287:EST485 288:EST486 289:EST487 290:EST488 291:EST489 292:EST490 293:EST491 294:EST492 295:EST493 296:EST494 297:EST495 298:EST496 299:EST497 300:EST498 301:EST499 302:EST500 303:EST501 304:EST502 305:EST503 306:EST504 307:EST505 308:EST506 309:EST507 310:EST508 311:EST509 312:EST510 313:EST511 314:EST512 315:EST513 316:EST514 317:EST515 318:EST516 319:EST517 320:EST518 321:EST519 322:EST520 323:EST521 324:EST522 325:EST523 326:EST524 327:EST525 328:EST526 329:EST527 330:EST528 331:EST529 332:EST530 333:EST531 334:EST532 335:EST533 336:EST534 337:EST535 338:EST536 339:EST537 340:EST538 341:EST539 342:EST540 343:EST541 344:EST542 345:EST543 346:EST544 347:EST545 348:EST546 349:EST547 350:EST548 351:EST549 352:EST550 353:EST551 354:EST552 355:EST553 356:EST554 357:EST555 358:EST556 359:EST557 360:EST558 361:EST559 362:EST560 363:EST561 364:EST562 365:EST563 366:EST564 367:EST565 368:EST566 369:EST567 370:EST568 371:EST569 372:EST570 373:EST571 374:EST572 375:EST573 376:EST574 377:EST575 378:EST576 379:EST577 380:EST578 381:EST579 382:EST580 383:EST581 384:EST582 385:EST583 386:EST584 387:EST585 388:EST586 389:EST587 390:EST588 391:EST589 392:EST590 393:EST591 394:EST592 395:EST593 396:EST594 397:EST595 398:EST596 399:EST597 400:EST598 401:EST599 402:EST600 403:EST601 404:EST602 405:EST603 406:EST604 407:EST605 408:EST606 409:EST607 410:EST608 411:EST609 412:EST610 413:EST611 414:EST612 415:EST613 416:EST614 417:EST615 418:EST616 419:EST617 420:EST618 421:EST619 422:EST620 423:EST621 424:EST622 425:EST623 426:EST624 427:EST625 428:EST626 429:EST627 430:EST628 431:EST629 432:EST630 433:EST631 434:EST632 435:EST633 436:EST634 437:EST635 438:EST636 439:EST637 440:EST638 441:EST639 442:EST640 443:EST641 444:EST642 445:EST643 446:EST644 447:EST645 448:EST646 449:EST647 450:EST648 451:EST649 452:EST650 453:EST651 454:EST652 455:EST653 456:EST654 457:EST655 458:EST656 459:EST657 460:EST658 461:EST659 462:EST660 463:EST661 464:EST662 465:EST663 466:EST664 467:EST665 468:EST666 469:EST667 470:EST668 471:EST669 472:EST670 473:EST671 474:EST672 475:EST673 476:EST674 477:EST675 478:EST676 479:EST677 480:EST678 481:EST679 482:EST680 483:EST681 484:EST682 485:EST683 486:EST684 487:EST685 488:EST686 489:EST687 490:EST688 491:EST689 492:EST690 493:EST691 494:EST692 495:EST693 496:EST694 497:EST695 498:EST696 499:EST697 500:EST698 501:EST699 502:EST700 503:EST701 504:EST702 505:EST703 506:EST704 507:EST705 508:EST706 509:EST707 510:EST708 511:EST709 512:EST710 513:EST711 514:EST712 515:EST713 516:EST714 517:EST715 518:EST716 519:EST717 520:EST718 521:EST719 522:EST720 523:EST721 524:EST722 525:EST723 526:EST724 527:EST725 528:EST726 529:EST727 530:EST728 531:EST729 532:EST730 533:EST731 534:EST732 535:EST733 536:EST734 537:EST735 538:EST736 539:EST737 540:EST738 541:EST739 542:EST740 543:EST741 544:EST742 545:EST743 546:EST744 547:EST745 548:EST746 549:EST747 550:EST748 551:EST749 552:EST750 553:EST751 554:EST752 555:EST753 556:EST754 557:EST755 558:EST756 559:EST757 560:EST758 561:EST759 562:EST760 563:EST761 564:EST762 565:EST763 566:EST764 567:EST765 568:EST766 569:EST767 570:EST768 571:EST769 572:EST770 573:EST771 574:EST772 575:EST773 576:EST774 577:EST775 578:EST776 579:EST777 580:EST778 581:EST779 582:EST780 583:EST781 584:EST782 585:EST783 586:EST784 587:EST785 588:EST786 589:EST787 590:EST788 591:EST789 592:EST790 593:EST791 594:EST792 595:EST793 596:EST794 597:EST795 598:EST796 599:EST797 600:EST798 601:EST799 602:EST800 603:EST801 604:EST802 605:EST803 606:EST804 607:EST805 608:EST806 609:EST807 610:EST808 611:EST809 612:EST810 613:EST811 614:EST812 615:EST813 616:EST814 617:EST815 618:EST816 619:EST817 620:EST818 621:EST819 622:EST820 623:EST821 624:EST822 625:EST823 626:EST824 627:EST825 628:EST826 629:EST827 630:EST828 631:EST829 632:EST830 633:EST831 634:EST832 635:EST833 636:EST834 637:EST835 638:EST836 639:EST837 640:EST838 641:EST839 642:EST840 643:EST841 644:EST842 645:EST843 646:EST844 647:EST845 648:EST846 649:EST847 650:EST848 651:EST849 652:EST850 653:EST851 654:EST852 655:EST853 656:EST854 657:EST855 658:EST856 659:EST857 660:EST858 661:EST859 662:EST860 663:EST861 664:EST862 665:EST863 666:EST864 667:EST865 668:EST866 669:EST867 670:EST868 671:EST869 672:EST870 673:EST871 674:EST872 675:EST873 676:EST874 677:EST875 678:EST876 679:EST877 680:EST878 681:EST879 682:EST880 683:EST881 684:EST882 685:EST883 686:EST884 687:EST885 688:EST886 689:EST887 690:EST888 691:EST889 692:EST890 693:EST891 694:EST892 695:EST893 696:EST894 697:EST895 698:EST896 699:EST897 700:EST898 701:EST899 702:EST900 703:EST901 704:EST902 705:EST903 706:EST904 707:EST905 708:EST906 709:EST907 710:EST908 711:EST909 712:EST910 713:EST911 714:EST912 715:EST913 716:EST914 717:EST915 718:EST916 719:EST917 720:EST918 721:EST919 722:EST920 723:EST921 724:EST922 725:EST923 726:EST924 727:EST925 728:EST926 729:EST927 730:EST928 731:EST929 732:EST930 733:EST931 734:EST932 735:EST933 736:EST934 737:EST935 738:EST936 739:EST937 740:EST938 741:EST939 742:EST940 743:EST941 744:EST942 745:EST943 746:EST944 747:EST945 748:EST946 749:EST947 750:EST948 751:EST949 752:EST950 753:EST951 754:EST952 755:EST953 756:EST954 757:EST955 758:EST956 759:EST957 760:EST958 761:EST959 762:EST960 763:EST961 764:EST962 765:EST963 766:EST964 767:EST965 768:EST966 769:EST967 770:EST968 771:EST969 772:EST970 773:EST971 774:EST972 775:EST973 776:EST974 777:EST975 778:EST976 779:EST977 780:EST978 781:EST979 782:EST980 783:EST981 784:EST982 785:EST983 786:EST984 787:EST985 788:EST986 789:EST987 790:EST988 791:EST989 792:EST990 793:EST991 794:EST992 795:EST993 796:EST994 797:EST995 798:EST996 799:EST997 800:EST998 801:EST999 802:EST1000 803:EST1001 804:EST1002 805:EST1003 806:EST1004 807:EST1005 808:EST1006 809:EST1007 810:EST1008 811:EST1009 812:EST1010 813:EST1011 814:EST1012 815:EST1013 816:EST1014 817:EST1015 818:EST1016 819:EST1017 820:EST1018 821:EST1019 822:EST1020 823:EST1021 824:EST1022 825:EST1023 826:EST1024 827:EST1025 828:EST1026 829:EST1027 830:EST1028 831:EST1029 832:EST1030 833:EST1031 834:EST1032 835:EST1033 836:EST1034 837:EST1035 838:EST1036 839:EST1037 840:EST1038 841:EST1039 842:EST1040 843:EST1041 844:EST1042 845:EST1043 846:EST1044 847:EST1045 848:EST1046 849:EST1047 850:EST1048 851:EST1049 852:EST1050 853:EST1051 854:EST1052 855:EST1053 856:EST1054 857:EST1055 858:EST1056 859:EST1057 860:EST1058 861:EST1059 862:EST1060 863:EST1061 864:EST1062 865:EST1063 866:EST1064 867:EST1065 868:EST1066 869:EST1067 870:EST1068 871:EST1069 872:EST1070 873:EST1071 874:EST1072 875:EST1073 876:EST1074 877:EST1075 878:EST1076 879:EST1077 880:EST1078 881:EST1079 882:EST1080 883:EST1081 884:EST1082 885:EST1083 886:EST1084 887:EST1085 888:EST1086 889:EST1087 890:EST1088 891:EST1089 892:EST1090 893:EST1091 894:EST1092 895:EST1093 896:EST1094 897:EST1095 898:EST1096 899:EST1097 900:EST1098 901:EST1099 902:EST1100 903:EST1101 904:EST1102 905:EST1103 906:EST1104 907:EST1105 908:EST1106 909:EST1107 910:EST1108 911:EST1109 912:EST1110 913:EST1111 914:EST1112 915:EST1113 916:EST1114 917:EST1115 918:EST1116 919:EST1117 920:EST1118 921:EST1119 922:EST1120 923:EST1121 924:EST1122 925:EST1123 926:EST1124 927:EST1125 928:EST1126 929:EST1127 930:EST1128 931:EST1129 932:EST1130 933:EST1131 934:EST1132 935:EST1133 936:EST1134 937:EST1135 938:EST1136 939:EST1137 940:EST1138 941:EST1139 942:EST1140 943:EST1141 944:EST1142 945:EST1143 946:EST1144 947:EST1145 948:EST1146 949:EST1147 950:EST1148 951:EST1149 952:EST1150 953:EST1151 954:EST1152 955:EST1153 956:EST1154 957:EST1155 958:EST1156 959:EST1157 960:EST1158 961:EST1159 962:EST1160 963:EST1161 964:EST1162 965:EST1163 966:EST1164 967:EST1165 968:EST1166 969:EST1167 970:EST1168 971:EST1169 972:EST1170 973:EST1171 974:EST1172 975:EST1173 976:EST1174 977:EST1175 978:EST1176 979:EST1177 980:EST1178 981:EST1179 982:EST1180 983:EST1181 984:EST1182 985:EST1183 986:EST1184 987:EST1185 988:EST1186 989:EST1187 990:EST1188 991:EST1189 992:EST1190 993:EST1191 994:EST1192 995:EST1193 996:EST1194 997:EST1195 998:EST1196 999:EST1197 1000:EST1198 1001:EST1199 1002:EST1200 1003:EST1201 1004:EST1202 1005:EST1203 1006:EST1204 1007:EST1205 1008:EST1206 1009:EST1207 1010:EST1208 1011:EST1209 1012:EST1210 1013:EST1211 1014:EST1212 1015:EST1213 1016:EST1214 1017:EST1215 1018:EST1216 1019:EST1217 1020:EST1218 1021:EST1219 1022:EST1220 1023:EST1221 1024:EST1222 1025:EST1223 1026:EST1224 1027:EST1225 1028:EST1226 1029:EST1227 1030:EST1228 1031:EST1229 1032:EST1230 1033:EST1231 1034:EST1232 1035:EST1233 1036:EST1234 1037:EST1235 1038:EST1236 1039:EST1237 1040:EST1238 1041:EST1239 1042:EST1240 1043:EST1241 1044:EST1242 1045:EST1243 1046:EST1244 1047:EST1245 1048:EST1246 1049:EST1247 1050:EST1248 1051:EST1249 1052:EST1250 1053:EST1251 1054:EST1252 1055:EST1253 1056:EST1254 1057:EST1255 1058:EST1256 1059:EST1257 1060:EST1258 1061:EST1259 1062:EST1260 1063:EST1261 1064:EST1262 1065:EST1263 1066:EST1264 1067:EST1265 1068:EST1266 1069:EST1267 1070:EST1268 1071:EST1269 1072:EST1270 1073:EST1271 1074:EST1272 1075:EST1273 1076:EST1274 1077:EST1275 1078:EST1276 1079:EST1277 1080:EST1278 1081:EST1279 1082:EST1280 1083:EST1281 1084:EST1282 1085:EST1283 1086:EST1284 1087:EST1285 1088:EST1286 1089:EST1287 1090:EST1288 1091:EST1289 1092:EST1290 1093:EST1291 1094:EST1292 1095:EST1293 1096:EST1294 1097:EST1295 1098:EST1296 1099:EST1297 1100:EST1298 1101:EST1299 1102:EST1300 1103:EST1301 1104:EST1302 1105:EST1303 1106:EST1304 1107:EST1305 1108:EST1306 1109:EST1307 1110:EST1308 1111:EST1309 1112:EST1310 1113:EST1311 1114:EST1312 1115:EST1313 1116:EST1314 1117:EST1315 1118:EST1316 1119:EST1317 1120:EST1318 1121:EST1319 1122:EST1320 1123:EST1321 1124:EST1322 1125:EST1323 1126:EST1324 1127:EST1325 1128:EST1326 1129:EST1327 1130:EST1328 1131:EST1329 1132:EST1330 1133:EST1331 1134:EST1332 1135:EST1333 1136:EST1334 1137:EST1335 1138:EST1336 1139:EST1337 1140:EST1338 1141:EST1339 1142:EST1340 1143:EST1341 1144:EST1342 1145:EST1343 1146:EST1344 1147:EST1345 1148:EST1346 1149:EST1347 1150:EST1348 1151:EST1349 1152:EST1350 1153:EST1351 1154:EST1352 1155:EST1353 1156:EST1354 1157:EST1355 1158:EST1356 1159:EST1357 1160:EST1358 1161:EST1359 1162:EST1360 1163:EST1361 1164:EST1362 1165:EST1363 1166:EST1364 1167:EST1365 1168:EST1366 1169:EST1367 1170:EST1368 1171:EST1369 1172:EST1370 1173:EST1371 1174:EST1372 1175:EST1373 1176:EST1374 1177:EST1375 1178:EST1376 1179:EST1377 1180:EST1378 1181:EST1379 1182:EST1380 1183:EST1381 1184:EST1382 1185:EST1383 1186:EST1384 1187:EST1385 1188:EST1386 1189:EST1387 1190:EST1388 1191:EST1389 1192:EST1390 1193:EST1391 1194:EST1392 1195:EST1393 1196:EST1394 1197:EST1395 1198:EST1396 1199:EST1397 1200:EST1398 1201:EST1399 1202:EST1400 1203:EST1401 1204:EST1402 1205:EST1403 1206:EST1404 1207:EST1405 1208:EST1406 1209:EST1407 1210:EST1408 1211:EST1409 1212:EST1410 1213:EST1411 1214:EST1412 1215:EST1413 1216:EST1414 1217:EST1415 1218:EST1416 1219:EST1417 1220:EST1418 1221:EST1419 1222:EST1420 1223:EST1421 1224:EST1422 1225:EST1423 1226:EST1424 1227:EST1425 1228:EST1426 1229:EST1427 1230:EST1428 1231:EST1429 1232:EST143

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -28ml3 rev2 ET from Amersham  
High quality sequence stop: 448.

# FEATURES

Source  
1..463  
/organism="Homo sapiens"  
/note="Organ: mixed (see below); Vector: pT73D-Pac  
(Pharmacia) with a modified polylinker; Site\_1: Not 1;  
Site\_2: Eco RI; Equal amounts of plasmid DNA from three  
normalized libraries (melanocyte 2NbHM, pregnant uterus  
NbHPU, and fetal heart NbH19W) were mixed and ss circles  
were made in vitro. Following HAP purification, this DNA  
was used as tracer in a subtractive hybridization  
reaction. The driver was PCR-amplified cDNAs from pools of  
5,000 clones made from the same 3 libraries. The pools  
consisted of I.M.A.G.E. clones 260232-265223,  
340488-345479, and 484488-489479."  
/clone="666590"  
/clone\_lib="Soares NHMPu S1"  
/tissue\_type="pooled human melanocyte, fetal heart, and  
pregnant uterus"  
/lab\_host="DH10B"  
154 a 104 c 88 g 117 t  
<1..>463

# BASE COORDINATE

Query Match 21.5%; Score 255; DB 65; Length 463;  
Best Local Similarity 99.6%; Pred. No. 0.00e+00;  
Matches 256; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Db 1 aggcattgcaaacccagtcaccagatcctgaagctgctcagttttggcgaaataaaaaatg 60  
QY 929 AGGCAIGCAAAACCCAGTGACAGATCCTGAAGCTGCTCAGTTTGTGGCGAATAAAAAATG 988  
Db 61 ggcacaaagacaccccttggaaggccctaatgcacgactaaagcactcaagacgtaccact 120  
QY 989 GCGACCAAGACACCTTGGAAGGGCCCTAATGCAGCACTAAAGCAGCTACCACT 1048  
Db 121 ttcccaaaactgtcactcagagttctaaagaagacattcaggttctctcagagttccacaa 180  
QY 1049 TTCCCAAAACTGTCACTCAGAGTCTAAAGAAGACCATCAGTCTCTTACAGCTTCACAA 1108  
181 tgcacaaattgtatcagaagttatttttagaaatgatagggttaaccaggtccaatcagtaa 240  
QY 1109 TGTACAAATTGTATCAGAAGTTATTTTAGAAATGATAGGTAAACAGGTCCCAATCAGTAA 1168  
Db 241 aaataagctgcttataa 257  
QY 1169 AAATAAGCTGCTTATAA 1185

# RESULT

2  
ID HS1151638 standard; RNA; EST: 463 BP.  
AC AA233719;  
NI 91856711  
DT 06-MAR-1997 (Rel. 51, Created)  
DT 06-MAR-1997 (Rel. 51, Last updated, Version 1)  
DE z474908.r1 Soares NHMPu S1 Homo sapiens cDNA clone 666590 5'.  
KW EST.  
OS Homo sapiens (human)  
OC Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;  
OC Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP 1-463  
RA Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M.,  
RA Holman M., Hultman M., Kucaba T., Le M., Lennon G., Marra M.,  
RA Parsons J., Rifkin L., Rohlfing T., Soares M., Tan F.,  
RA Trevaskis E., Waterston R., Williamson A., Wohlmann P., Wilson R.;

"The WashU-Merck EST Project";  
Unpublished.

Contact: Wilson RK WashU-Merck EST Project Washington University  
School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis,  
MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email:

est@watson.wustl.edu This clone is available royalty-free through  
LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for  
further information. Seq primer: -28ml3 rev2 ET from Amersham High  
quality sequence stop: 448.

Location/Qualifiers

Key

1..463

/organism="Homo sapiens"

/note="Organ: mixed (see below); Vector: pT73D-Pac  
(Pharmacia) with a modified polylinker; Site\_1: Not 1;  
Site\_2: Eco RI; Equal amounts of plasmid DNA from three  
normalized libraries (melanocyte 2NbHM, pregnant uterus  
NbHPU, and fetal heart NbH19W) were mixed, and ss circles  
were made in vitro. Following HAP purification, this DNA  
was used as tracer in a subtractive hybridization reaction

The driver was PCR-amplified cDNAs from pools of 5,000  
clones made from the same 3 libraries. The pools consisted  
of I.M.A.G.E. clones 260232-265223, 340488-345479, and  
484488-489479."

/clone="666590"

/clone\_lib="Soares NHMPu S1"

/tissue\_type="pooled human melanocyte, fetal heart, and

pregnant uterus"

/lab\_host="DH10B"

<1..>463

Sequence 463 BP; 154 A; 104 C; 88 G; 117 T; 0 Other;

Query Match 21.5%; Score 255; DB 83; Length 463;

Best Local Similarity 99.6%; Pred. No. 0.00e+00;

Matches 256; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 aggcattgcaaacccagtcaccagatcctgaagctgctcagttttggcgaaataaaaaatg 60

QY 929 AGGCAIGCAAAACCCAGTGACAGATCCTGAAGCTGCTCAGTTTGTGGCGAATAAAAAATG 988

Db 61 ggcacaaagacaccccttggaaggccctaatgcacgactaaagcactcaagacgtaccact 120

QY 989 GCGACCAAGACACCTTGGAAGGGCCCTAATGCAGCACTAAAGCAGCTACCACT 1048

Db 121 ttcccaaaactgtcactcagagttctaaagaagacattcaggttctctcagagttccacaa 180

QY 1049 TTCCCAAAACTGTCACTCAGAGTCTAAAGAAGACCATCAGTCTCTTACAGCTTCACAA 1108

Db 181 tgcacaaattgtatcagaagttatttttagaaatgatagggttaaccaggtccaatcagtaa 240

QY 1109 TGTACAAATTGTATCAGAAGTTATTTTAGAAATGATAGGTAAACAGGTCCCAATCAGTAA 1168

Db 241 aaataagctgcttataa 257

QY 1169 AAATAAGCTGCTTATAA 1185

RESULT 3

LOCUS AA195113 530 bp mRNA

DEFINITION zr35a03.r1 Soares NHMPu S1 Homo sapiens cDNA clone 665356 5'.

ACCESSION AA195113

NID 91784803

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;

Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 530)

AUTHORS

Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M.,  
Holman M., Hultman M., Kucaba T., Le M., Lennon G., Marra M.,  
Parsons J., Rifkin L., Rohlfing T., Soares M., Tan F.,  
Trevaskis E., Waterston R., Williamson A., Wohlmann P. and







Query Match	1.9%; Score 22; DB 48; Length 450;	
Best Local Similarity	63.6%; Pred. No. 2.15e-04;	
Matches	35; Conservative 0; Mismatches 20; Indels 0; Gaps 0;	
Db	372 acagagagatnagngagagaaacannanagataggaataaaagcatgcaaa 426	
QY	885 ACCGGGAAGAAATGGGACGAGAGACATGTGAAACAAATAAAGGATGCAAA 939	
RESULT	8	
ID	AT4774 standard; RNA; EST; 472 BP.	
AC	T42477	
NI	G933235	
DT	02-FEB-1995 (Rel. 42, Created)	
DE	12-MAR-1997 (Rel. 51, Last updated, Version 15)	
DE	5740 Arabidopsis thaliana cDNA clone 113K157.	
KW	EST.	
OS	Arabidopsis thaliana (thale cress)	
OC	Eukaryotes; mitochondrial eukaryotes; Viridiplantae;	
OC	Charophyta; Embryophyta group; Embryophyta; Magnoliophyta;	
OC	Magnoliopsida; Capparales; Brassicaceae; Arabidopsis.	
OC	[1]	
OC	1-472	
OC	MEDLINE: 95148729.	
OC	Newman T., deBruin F.J., Green P., Keestra K., Kende H.,	
OC	McIntosh L., Ohlrogge J., Raikhel N., Somerville S., Thomashow M.,	
OC	Retzel E., Somerville C.;	
OC	"Genes galore: a summary of methods for accessing results from	
OC	large-scale partial sequencing of anonymous Arabidopsis cDNA	
OC	clones";	
OC	Plant Physiol. 106:1241-1255(1994).	
OC	AGIS: T42477; AGIS July 1995.	
OC	Contact: Thomas Newman MSU-DOE Plant Research Laboratory Michigan	
OC	State University MSU-DOE-PRU, Michigan State University, Plant	
OC	Biology Bldg., E. Lansing, MI Tel: 517-353-0854 Fax: 517-353-9168	
OC	Email: 22313tcm@bm.cl.msu.edu. NCBI gi: 933235	
OC	Key Location/Qualifiers	
OC	Source	
OC	1..472	
OC	/organism="Arabidopsis thaliana"	
OC	/clone="113K157"	
OC	/strain="var columbia"	
OC	/note="thale cress"	
OC	Sequence 472 BP; 126 A; 97 C; 100 G; 137 T; 12 other;	
Query Match	1.9%; Score 22; DB 77; Length 472;	
Best Local Similarity	68.6%; Pred. No. 2.15e-04;	
Matches	35; Conservative 0; Mismatches 16; Indels 0; Gaps 0;	
Db	365 gaaacagtgtcgtgaagctacattgagttggtcccaagaacacctnag 415	
QY	64 GAAACGTTCTTCCCAAGTACCTTATTATGACGAAGAACCTCTCATCAG 114	
RESULT	9	
ID	WM1135 standard; RNA; EST; 105 BP.	
AC	W10113;	
NI	G1284430	
DT	29-APR-1996 (Rel. 47, Created)	
DE	05-MAR-1997 (Rel. 51, Last updated, Version 2)	
DE	ma42d03.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone 313349	
DE	5,	
KW	EST.	
OS	Mus musculus (house mouse)	
OC	Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;	
OC	Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;	
OC	Mus.	
OC	[1]	
OC	1-105	
OC	Marra M., Hillier L., Allen M., Bowles M., Dietrich N.,	
OC	Dubuque T., Geisel S., Kucaba T., Lacy M., Martin J.,	
OC	Morris M., Schellenberg K., Steptoe M., Tan F., Underwood K.,	
OC	Moore B., Theising B., Wylie T., Lennon G., Soares B., Wilson R.,	



```

/tissue_type="Thymus"
/dev_stage="4 weeks"
/lab_host="DH10B"
<1..>107
MRNA 50 a 16 c 24 g 17 t
BASE COUNT
ORIGIN
Query Match 1.8%; Score 21; DB 58; Length 107;
Best Local Similarity 76.9%; Pred. No. 3.88e-03;
Matches 30; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
Db 65 aaagaaatgggaaaaaagaaaaaagaaaaaagaaaaaagaaaaa 103
||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 891 AAAGAAAGTGGGACAGACATTGAAAAAACAATAAA 929

RESULT 13
LOCUS HUMUT5145 194 bp DNA STS 28-MAY-1993
DEFINITION Human STS UT5145.
ACCESSION L17708
KEYWORDS
SOURCE Homo sapiens
ORGANISM
PCR primer: STS sequence; microsatellite marker; microsatellite
repeat; repeat polymorphism; sequence tagged site.
Homo sapiens DNA.
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 194)
Gerken,S.C., Matsunami,N., Lawrence,E., Carlson,M., Moore,M.,
Ballard,L., Melis,R., Robertson,M., Bradley,P., Elsner,T.,
Tingey,A., Rodriguez,P., Albertsen,H., Lalouel,J.-M. and White,R.
Genetic and physical mapping of simple sequence repeat containing
sequence tagged sites from the human genome
Unpublished (1993)
Submitted by: Utah Center for Human Genome Research University of
Utah, Dept. of Human Genetics
2160 Eccles Institute of Human Genetics
Salt Lake City, UT 84112
e-mail: sts@corona.med.utah.edu
Primer A: CTACTCAGGAGGACGAC
Primer B: ACCAAATCCACTTCACCTAGA
32P-label: A Primer
PCR Profile:
Initial Denaturation: 94C 300sec
PCR Cycles: 5
Denaturation: 94C 10sec
Annealing: 58C 10sec
Extension: 72C 20sec
Mg++: 1mM
Gel: Acrylamide 7%, Formamide 32%, Urea 34%
Alleles: 1
FEATURES
Location/Qualifiers
source 1..194
STS /organism="Homo sapiens"
5..181
/standard_names="STS UT5145"
primer_bind 5..22
primer_bind complement(161..181)
BASE COUNT 108 a 11 c 55 g 16 t 4 others
ORIGIN
Query Match 1.8%; Score 21; DB 48; Length 194;
Best Local Similarity 70.0%; Pred. No. 3.88e-03;
Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
Db 61 ggaagaaagaagaagaaagaaagaaagaaagaaagaaagaaagaa 110
||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 889 GGAAGAAAGTGGACACAGACATTGAAAAAACAATAAAGGCATGCA 938

RESULT 14
LOCUS G11923 245 bp DNA STS 23-OCT-1995
DEFINITION human STS MR4116.

```

```

ACCESSION G11923
NID g1036742
KEYWORDS STS sequence; primer; sequence tagged site.
SOURCE human STS derived from random genomic DNA.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Chonata;
Catarrhini; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates;
1 (bases 1 to 245)
Hudson,T.
Whitehead Institute/MIT Center for Genome Research; Physically
Mapped STS
Unpublished (1995)

```

```

Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Whitehead Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu
Primer A: TGTGTGTAGTTGTTGTTGTTATTGGA
Primer B: AAAGGAGTCAAAATGGGTTTT
STS size: 100
PCR Profile:
Presoak:
Annealing: 56 degrees C
Polymerization:
PCR Cycles: 35
Thermal Cycler:
Protocol:
Template: 10 ng
Primer: each 5 pM
dNTPs: each 4 nM
Taq Polymerase: 0.025 units/ul
Total Vol: 20 ul
Buffer:
MgCl2: 1.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 9.3

```

Prepared with primer pairs derived from random genomic sequence.

```

FEATURES
Location/Qualifiers
source 1..245
STS /organism="Homo sapiens"
56..155
/map="791_B_4; 816_F_6; 921_C_9; 928_A_7; 934_F_6;
963_G_6"
56..80
/map="791_B_4; 816_F_6; 921_C_9; 928_A_7; 934_F_6;
963_G_6"
primer_bind complement(134..155)
primer_bind /map="791_B_4; 816_F_6; 921_C_9; 928_A_7; 934_F_6;
963_G_6"
BASE COUNT 92 a 28 c 35 g 87 t 3 others
ORIGIN
Query Match 1.8%; Score 21; DB 40; Length 245;
Best Local Similarity 70.8%; Pred. No. 3.88e-03;
Matches 34; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
Db 109 tncctgaatgttacctcatttaaaaaaacccatttgactcctttt 156
||||| ||||| ||||| ||||| ||||| ||||| |||||
CP 453 TGCATTAGACGTCTCAITTTGAGAAGAACCCATCTGGACATCTTT 406

RESULT 15
LOCUS G06067 253 bp DNA STS 19-OCT-1995

```

Search completed: Tue Dec 2 16:04:01 1997  
Job time : 747 secs.

DEFINITION human STS WI-6406.  
ACCESSION G06067  
NID g859312  
KEYWORDS STS sequence; primer; sequence tagged site.  
SOURCE human STSS derived from sequences in dbEST and the Unigene collection.

ORGANISM Homo sapiens  
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 253)  
Hudson, T.  
Whitehead Institute/MIT Center for Genome Research; Physically Mapped ESTs  
Unpublished (1995)

CONTACT: Thomas Hudson  
Whitehead Institute/MIT Center for Genome Research  
Whitehead Institute for Biomedical Research  
9 Cambridge Center, Cambridge MA 02142 USA  
Tel: 617 252 1900  
Fax: 617 252 1902  
Email: thudson@genome.wi.mit.edu

Primer A: CATTGCTTTTCATACATACAGTTC  
Primer B: CACAGTTGTGAAGCGTGC  
STS size: 225  
PCR Profile:  
  Prisoak:  
    Denaturation:  
    Annealing: 56 degrees C  
    Polymerization:  
    PCR Cycles: 35  
    Thermal Cycler:  
Protocol:  
  Template: 10 ng  
  Primer: each 5 pM  
  dNTPs: each 4 nM  
  Taq Polymerase: 0.025 units/ul  
  Total Vol: 20 ul

Buffer:  
  MgCl2: 1.5 mM  
  KCl: 50 mM  
  Tris-HCl: 10 mM  
  pH: 9.3

Prepared with primer pairs derived from 238433 -- dbEST.  
Location/Qualifiers  
1..253  
  /organism="Homo sapiens"  
  /map="923\_H\_8; 959\_F\_5; 441.7 cR from top of Chr17 linkage group"  
primer\_bind 29..253  
  /map="923\_H\_8; 959\_F\_5; 441.7 cR from top of Chr17 linkage group"  
primer\_bind complement(235..253)  
  /map="923\_H\_8; 959\_F\_5; 441.7 cR from top of Chr17 linkage group"  
BASE COUNT 76 a 57 c 49 g 70 t 1 others  
ORIGIN

Query Match 1.8%; Score 21; DB 38; Length 253;  
Best Local Similarity 78.4%; Pred. No. 3.88e-03;  
Matches 29; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Db 123 acacagctcacatgtacagacaataaactgctcaag 159  
||||| | | | | | | | | | | | | | | | | | | | |  
Oy 708 ACACAGCTCACAGACAGACAGACTTCCAGCTGCTGAAG 744

\*\*\*\*\*



(TM)

\*\*\*\*\*

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

MParch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 2 16:04:23 1997; MasPar time 910.66 Seconds  
1313.360 Million cell updates/sec

ular output not generated.

Title: >US-08-915-004-10  
Description: (1-1089) from US08915004.seq  
Perfect Score: 1089

N.A. Sequence: 1 ATGAACAAGTTGCTGCTGCTG.....TAAATAAGCTGCTTATAA 1089  
Comp: TACTTGTTCACGACACGAC.....ATTTTATTCGACGAATTT

Scoring table: TABLE default  
Gap 6

Nmatch STD : Dbase 0; Query 0  
Searched: 362067 seqs, 549138275 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: emb1-new3  
1: BCT 2: FUN 3: GEN1 4: GEN2 5: HTG1 6: HTG2 7: HUM 8: INV  
9: ORG 10: MAM 11: VRT 12: PLN 13: PRO 14: ROD 15: SYN 16: UNC  
17: VIR

Database: genbank99  
18: BCT1 19: BCT2 20: BCT3 21: BCT4 22: BCT5 23: BCT6 24: BCT7  
25: BCT8 26: BCT9 27: BCT10 28: BCT11 29: GEN1 30: GEN2  
31: GEN3 32: HTG1 33: HTG2 34: HTG3 35: INV1 36: INV2 37: INV3  
38: INV4 39: INV5 40: INV6 41: INV7 42: INV8 43: INV9 44: INV10  
45: INV11 46: MAM1 47: MAM2 48: MAM3 49: VRT1 50: VRT2 51: VRT3  
52: VRT4 53: PAT1 54: PAT2 55: PAT3 56: PAT4 57: PAT5 58: PHG  
59: PLN1 60: PLN2 61: PLN3 62: PLN4 63: PLN5 64: PLN6 65: PLN7  
66: PLN8 67: PLN9 68: PLN10 69: PLN11 70: PRI1 71: PRI2  
72: PRI3 73: PRI4 74: PRI5 75: PRI6 76: PRI7 77: PRI8 78: PRI9  
79: PRI10 80: PRI11 81: PRI12 82: PRI13 83: PRI14 84: PRI15  
85: ROD1 86: ROD2 87: ROD3 88: ROD4 89: ROD5 90: ROD6 91: ROD7  
92: ROD8 93: STR 94: SYN 95: UNA 96: VRL1 97: VRL2 98: VRL3  
99: VRL4 100: VRL5 101: VRL6 102: VRL7 103: VRL8 104: VRL9  
105: VRL10  
genbank-new3  
106: BCT 107: GEN1 108: GEN2 109: HTG1 110: HTG2 111: INV  
112: MAM 113: VRT 114: PHG 115: PLN 116: PRI1 117: PRI2  
118: ROD 119: SYN 120: UNA 121: VRL  
Database: u-emb150\_99  
122: part1

Statistics: Mean 11.131; Variance 4.282; scale 2.599

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	30	2.8	215	57	Sequence 5 from paten	4.40e-05
2	28	2.6	215	57	Sequence 5 from paten	1.20e-03
3	28	2.6	354	111	Oxytricha fallax 57kD	1.20e-03
4	28	2.6	354	8	Oxytricha fallax 57kD	1.20e-03
5	25	2.3	354	111	Oxytricha fallax 57kD	1.33e-01
6	25	2.3	354	8	Oxytricha fallax 57kD	1.33e-01
7	24	2.2	7095	92	R.norvegicus mRNA for	5.89e-01
8	24	2.2	115419	34	Human DNA sequence **	5.89e-01
9	23	2.1	1851	117	H.sapiens LIPA gene,	2.50e+00
10	23	2.1	1851	75	H.sapiens LIPA gene,	2.50e+00
11	23	2.1	3021	69	S.cerevisiae nitrogen	2.50e+00
12	23	2.1	3177	43	L.stagnalis mRNA for	2.50e+00
13	23	2.1	38586	37	Caenorhabditis elegans	2.50e+00
14	23	2.1	43100	65	Saccharomyces cerevis	2.50e+00
15	23	2.1	145750	33	Human DNA sequence **	2.50e+00
16	22	2.0	370	8	Oxytricha fallax 57kD	1.01e+01
17	22	2.0	370	111	Oxytricha fallax 57kD	1.01e+01
18	22	2.0	1245	61	C.reinhardtii phospho	1.01e+01
19	22	2.0	1256	47	O.cuniculus SPL7 gene	1.01e+01
20	22	2.0	1380	91	R.norvegicus interieu	1.01e+01
21	22	2.0	1388	86	M.musculus tumor necr	1.01e+01
22	22	2.0	1388	118	M.musculus tumor necr	1.01e+01
23	22	2.0	1505	89	M.musculus tumor necr	1.01e+01
24	22	2.0	2492	65	S.cerevisiae fus2 gen	1.01e+01
25	22	2.0	2552	55	Sequence 5 from Paten	1.01e+01
26	22	2.0	2787	91	Rat peroxisomal 3-ket	1.01e+01
27	22	2.0	3112	81	Human LD78 beta gene.	1.01e+01
28	22	2.0	3796	90	Murine tumor necrosis	1.01e+01
29	22	2.0	4561	38	Caenorhabditis elegans	1.01e+01
30	22	2.0	7857	88	Mouse mRNA for mouse	1.01e+01
31	22	2.0	8018	86	Mouse MAP1B mRNA for	1.01e+01
32	22	2.0	9416	83	Human microtubule-ass	1.01e+01
33	22	2.0	17013	20	E.amylovora (Ea7/74)	1.01e+01
34	22	2.0	26923	38	Caenorhabditis elegans	1.01e+01
35	22	2.0	28687	37	C.elegans cosmid f42	1.01e+01
36	22	2.0	40397	65	S.cerevisiae chromoso	1.01e+01
37	22	2.0	90468	34	Human DNA sequence **	1.01e+01
38	22	2.0	185775	80	Homo sapiens fragile	1.01e+01
39	22	2.0	257912	103	Paramoecium bursaria	1.01e+01
40	21	1.9	814	105	Vesicular stomatitis	3.85e+01
41	21	1.9	814	104	RHVSVIN	3.85e+01
42	21	1.9	1451	45	T.cruzi kinetoplast m	3.85e+01
43	21	1.9	37844	110	Caenorhabditis elegans	3.85e+01
44	21	1.9	112762	117	Human DNA sequence fr	3.85e+01
45	21	1.9	201239	109	Human DNA sequence **	3.85e+01

ALIGNMENTS

RESULT 1 I28278 215 bp DNA PAT 30-OCT-1996  
LOCUS Sequence 5 from patent US 5569830.  
DEFINITION I28278  
ACCESSION 5569830  
NID g1819054  
KEYWORDS Unknown.  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 215)  
AUTHORS Bennett A., Labavitch J.M., Powell A. and Stotz H.  
TITLE Plant inhibitors of fungal polygalacturonases and their use to control fungal disease  
JOURNAL Patent: US 5569830-A 5 29-OCT-1996;  
FEATURES Location/Qualifiers  
source 1..215  
BASE COUNT 15 a 8 c 25 g 26 t 141 others  
ORIGIN

```

Query Match      2.8%; Score 30; DB 57; Length 215;
Best Local Similarity 13.3%; Pred. No. 4.40e-05;
Matches          26; Conservative 80; Mismatches 86; Indels 3; Gaps 3;

Db 7 sssvsvrtasndakkdngntsswttdccnrttgvvcdtdtttyrvnndshnkyssany 66
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 330 GGCCTACCTTGAGATAGAGTTCCTGTTGAAACATAGGAGCTGCCCTCTCGATTGGAGT 389
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 67 nyggnnvgaakthyythnvsdgadsktvdynasgtssnggtgdgnrsgadsvgsakta 126
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 390 GTGCAAGCTGAACCC-CAG-AGCGAATACAGTTTGCAAAAGATGCCAGATGGGTTC 447

Db 127 mtsrntgktannavdrnmgsdsvgsdntkthknsadgkvsknngdnrrnrygtgk 186
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 448 TTCTCAATGAGACGCTCATCTAA-AGCACCCCTGTAGAAAACACACAAATTGCAGTGTCT 506
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 187 snvsnnccgggnkrdv 201
    : | | | | : : : :
507 TGGTCTCTCGCTAAC 521

RESULT
LOCUS      2 128278 215 bp DNA PAT 30-OCT-1996
DEFINITION Sequence 5 from patent US 5569830.
ACCESSION 128278
NID       91819054
KEYWORDS  . Unknown.
SOURCE    Unknown.
ORGANISM  Unclassified.
REFERENCE 1 (bases 1 to 215)
AUTHORS  Bennett,A., Labavitch,J.M., Powell,A. and Stotz,H.
TITLE    Plant inhibitors of fungal polygalacturonases and their use to
JOURNAL  control fungal disease
          Patent: US 5569830-A 5 29-OCT-1996;

FEATURES
  source
    BASE COUNT  15 a  8 c  25 g  26 t  141 others
    ORIGIN
      Query Match      2.6%; Score 28; DB 57; Length 215;
      Best Local Similarity 17.0%; Pred. No. 1.20e-03;
      Matches          23; Conservative 55; Mismatches 55; Indels 2; Gaps 2;

39  rtwgvcdtdtttyrvnndshnkyssanynggnnygaakthyythnvsdgadsktvtddy 98
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
319 ATTGCGACACGCGGTGTGGTGGCGATTGCACTCTGCTTGACGCTAGCTGCAGCTCCTTGC 260

Db 99 n-asgtssngtgdgnrsgadsvgsdsktamttrntgktannavdrnmgsdsvgsdnt 157
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Cp 259 ACACGGGGCTGCAGTATAGACACTGCTCACTGGTGTGCCAGCTGTCTGTAGTAGTGGT 200

Db 158 kkhknsadgkvysk 172
    : : | : | : | :
Cp 199 C-AGGCAAGGGGCG 186

RESULT
LOCUS      3 0F089259 354 bp DNA INV 14-MAR-1997
DEFINITION Oxytricha fallax 57kD zinc finger/protein chimera gene, partial
ACCESSION 089259
NID       91881675
KEYWORDS  . Oxytricha fallax.
SOURCE    Oxytricha fallax.
ORGANISM  Eukaryotes; mitochondrial eukaryotes; Alveolata; Ciliophora;
          hypotrichs; Stichotrichida; Oxytricha.
REFERENCE 1 (bases 1 to 354)
AUTHORS  Doak,T.G., Doerder,F.P., Jahn,C.L. and Herrick,G.
TITLE    A proposed superfamily of transposase genes: transposon-like

elements in ciliated protozoa and a common 'D35E' motif"
Proc. Natl. Acad. Sci. U.S.A. 91:942-946(1994).

elements in ciliated protozoa and a common 'D35E' motif
Proc. Natl. Acad. Sci. U.S.A. 91 (3), 942-946 (1994)
94134747
2 (bases 1 to 354)
Witherspoon,D.J., Doak,T.G., Williams,K., Seger,J. and Herrick,G.
Selection on the protein-coding genes of the TBEL family of
transposable elements in the ciliates Oxytricha fallax and O.
trifallax
Unpublished
3 (bases 1 to 354)
Doak,T.G., Williams,K., Witherspoon,D.J. and Herrick,G.
Direct Submission
Submitted (11-FEB-1997) Oncological Science, University of Utah,
School of Med. Rm5C334, USA, UT 84132, USA
Location/Qualifiers
  1..354
  /organism="Oxytricha fallax"
  /strain="9D1"
  /transposon="TBEL"
  /note="this is a bulk sequence that was generated from a
PCR product that represents many transposon templates"
  <1..>354
  /codon_start=1
  /product="57kD zinc finger/protein chimera"
  /db_xref="FID:91881675"
  /transl_table=6
  /translation="HTRDLXKHLKAKHKKXXEXXXLKLXKRRKAREXXXXXXX
OAXEXXVXXRNXLSEXTKIMIKIQVKIPVLAQIDLTSLQSYLILEDSFDKVI
XDHQYAVKVPFNVL"
BASE COUNT  106 a  42 c  41 g  54 t  111 others
ORIGIN
  Query Match      2.6%; Score 28; DB 111; Length 354;
  Best Local Similarity 30.1%; Pred. No. 1.20e-03;
  Matches          25; Conservative 33; Mismatches 24; Indels 1; Gaps 1;

Db 110 hhydhvargwbgydgwgykaayaagcwgmgarwyaswtrtrytaargawdkrcqca 169
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 757 CTGCTGAAGTTATGGAACATCAAAACAAAGACCAAGA-TATAGTCAAGAAGATCATCCA 815
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 170 accwvytrrcmtcdgagrtmacd 192
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 816 AGATATTGACCTCTGTGAAACA 838

RESULT
ID      4 0F089259 standard; DNA; INV; 354 BP.
AC      089259;
NI      91881675
DT      13-MAR-1997 (Rel. 51, Created)
DT      13-MAR-1997 (Rel. 51, Last updated, Version 1)
DE      Oxytricha fallax 57kD zinc finger/protein chimera gene, partial
DE      cds.
KW      .
OS      Oxytricha fallax
OC      Eukaryotes; mitochondrial eukaryotes; Alveolata; Ciliophora;
OC      hypotrichs; Stichotrichida; Oxytrichidae; Oxytricha.
RN      [1]
RP      1-354
RX      MEDLINE; 94134747.
RA      Doak T.G., Doerder F.P., Jahn C.L., Herrick G.;
RT      "A proposed superfamily of transposase genes: transposon-like
RT      elements in ciliated protozoa and a common 'D35E' motif";
RL      Proc. Natl. Acad. Sci. U.S.A. 91:942-946(1994).
RN      [2]
RP      1-354
RA      Witherspoon D.J., Doak T.G., Williams K., Seger J., Herrick G.;
RT      "selection on the protein-coding genes of the TBEL family of
RT      transposable elements in the ciliates Oxytricha fallax and O.
RT      trifallax";
RL      Unpublished.
RN      [3]
RP      1-354

```











```

/codon_start=1
/db_xref="PID:g1086794"
/translation="MSEMERVLIINAVSRSRKREKWLPEEHVIVAGFLQKSSNP
SGSDVGGQKQVSGNSDKGEQSKGALLEKAKKTHKLVSRKKNHSCGFS
DVSPPNRSRSDISKSERASRSRSGGSDQLKQSKVPDNDSEVELLEIVAKR
TANQLIDPADSPDLVQSETHKCRVLPFRARLVFGSLIALTIACPGFITGLMNG
YSTVIGLYTFVSPKPKDOKDEWTFDDETHATTSPGLVDNIDQOGLGEGVI
YRGMNELRTYSPANYHNSQSVLRLEGMLRICRPAKAVLKHAFYDDPTLKQNO
PSMVSQVYDMKDAVSLRPRKARRWMSRKYPIHIFRAHSSSLVEIDHSSKAKSR
GMSRSANAPHNISETPGRHRSVDEETDGYTAESAEVEDEKRWKRVANSADIQ
EPRESNKKKGRSIIYLVRAARKERFWHLREACARNSPKVRKMSVIVKTCSS
ASLPEEIRGEDDATNTSVADALAAEKEDSESMFNSPKRICLPKEYLKYRSNTA
SPVKQIATILSIQVPRPKNSTVSDLTGMKAPGAQOISTELVDNINLATRIFD
FCRDMFVROVKQKQSLATIHPIFYIEKLELGELKFGTAAPKFTAVYIPKVDWMT
WDFEMKVGGLRLVLOTINLLKQSGSHOVKTEKRVNRWTSVRVTRSDSDLEP
PESSPDEFGAKNSEGTTKTKGKILSIVERAQQSLFQKAAKLAARLIEDYST
TPLMLNVEVEEGMTNIPPPSDRLMFAFRFRPIKLKRAVFGQGRSDVJSTVSE
WIETKROVLKLNLCVPMNDIILPVLUSGNPLHMGYNK"
complement(join(17297..17617,17660..17991,18105..18201,
18253..18396,18925..18948))
/evidence=not_experimental
/gene="F55C12.6"
/codon_start=1
/db_xref="PID:g1086795"
/translation="MATGIFRKIAMIKLKFQNHIIROISSKNKIISNLDXDVVIN
PDNLFEVMOGNGRDFDKEDYSTWSSWGDAYRFGILLKGTESIAYSFTIYQ
KSLGQDPDRHLGNAPVEPKYRGSILKVVTDYLIQBERMKHQNMLACNVHWSNFWL
KATGRSDIAGCTYVYISYVELYDKVPVPIKIDDCIEIVVKNVTKTVDVLKYDTSIFP
FNKRWIESLEFLEGIRVAYNSDGNVYGIGLCTYPSGECVISPFLYADNTKIAQKIFG
SMLEVIEKKNKIRFQVRNSDQCSNTE"
19575..19691
/evidence=not_experimental
/gene="F55C12.2"
/codon_start=1
/db_xref="PID:g1086796"
/translation="MRUGSSASSASVASRLYTRSRDSSLGSSSESDLIATFGDQD
ALNDFSOMNOLNFTKRYSELRNNSLDERTLKLTENSVLKRMHNEEOLTDNE
DRFKOLLSDERKARFTESRLRKELETSWNLYQWLEKOLISVKDAERSNEETK
RINLEKXENKLEAOLVVEGMEERITQERQPKKEAQQDIDSSSEWVEVLALE
TEELRKVDGPRSSISDRBMBHEIILKAKVAELKKEEMTDQLLASVERGRSL
IADTFLSDELAGDSQDLALRQEICNKLRYINGILMRVIERHPELEIGEEG
ILSKUTVRRISVLNVVVIIDSRVITDTFSKNLRLPSLIVSSGMNRRHMTSSTIG
TOFLTWHTATSLFSLFLPCDFQNALQMFCLTISFLVFPPTPSNIFELRATSQID
PVNG"
SE COUNT 12271 a 6976 c 6358 g 12981 t
ORIGIN

Query Match 2.1%; Score 23; DB 37; Length 38586;
Best Local Similarity 70.6%; Pred. No. 2.50e+00;
Matches 48; Conservative 0; Mismatches 19; Indels 1; Gaps 1;

Db 18273 aaagattatcaggattatcattacataataatcgccaaattgagaagaatttat 18332
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 978 AAGAAGACCATCAGGTTCTCTACAGCTTCAAT-GTACAAATGTATCAGAGATTAT 1036

Db 18333 ttctcga 18340
||||| |||||
Qy 1037 TTTTAGAA 1044

RESULT 14
LOCUS SCE9379 43100 bp DNA PLN 21-DEC-1994
DEFINITION Saccharomyces cerevisiae chromosome V cosmid 9379, 9581, and
lambda clone 4678.
ACCESSION U18796
NID 9603265

```

KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

GLN3 (positive nitrogen regulatory protein): MEI4 (meiosis-specific gene); HOM3 (aspartate kinase); CAU1 (dnaJ homolog); tRNA-Gln.  
baker's yeast strain-S288C (AB972).  
Saccharomyces cerevisiae  
Eukaryotae; mitochondrial eukaryotes; Fungi; Ascomycota;  
Hemiascomycetes; Saccharomycetales; Saccharomycetaceae;  
Saccharomycetes.  
1 (bases 1 to 43100)  
Dietrich,F.S.  
The sequence of S. cerevisiae cosmid 9379, 9581, and lambda clone 4678  
Unpublished (1994)  
2 (bases 1 to 43100)  
Dietrich,F.S., Mulligan,J., Hennesy,K., Allen,E., Araujo,R., Aviles,E., Berno,A., Brennan,T., Carpenter,J., Chen,E., Cherry,J.M., Chung,E., Duncan,M., Guzman,E., Hartzell,G., Hunkeler-Smith,S., Hyman,R., Kayser,A., Komp,C., Lashkari,D., Lew,H., Lin,D., Mosedale,D., Nakahara,K., Namath,A., Norgren,R., Oefner,P., Oh,C., Petel,F.X., Roberts,D., Sehl,P., Schramm,S., Shogren,T., Smith,V., Taylor,P., Wei,Y., Yelton,M., Botstein,D. and Davis,R.W.  
The complete sequence of Saccharomyces cerevisiae chromosome V  
Unpublished (1994)  
3 (bases 1 to 43100)  
Dietrich,F.S.  
Direct Submission  
Submitted (19-DEC-1994)  
Submitted by:  
Stanford DNA Sequence & Technology Center  
855 California Avenue  
Palo Alto, CA 94304, USA  
e-mail: dietrich@genome.stanford.edu

Neighboring Sequence:  
The 5' end of this sequence overlays with GenBank Accession Number U18778.

FEATURES  
source  
source  
source  
CDS

Location/Qualifiers  
1..43100  
/organism="Saccharomyces cerevisiae"  
/strain="S288C (AB972)"  
/chromosome="V"  
/map="5R"  
/strain="S288C (AB972)"  
/organism="Saccharomyces cerevisiae"  
/clone="cosmid 9581"  
/clone="cosmid 9379"  
/organism="Saccharomyces cerevisiae"  
complement(297..3527)  
/gene="YER033c"  
/codon\_start=1  
/evidence=not\_experimental  
/product="Yer033p"  
/db\_xref="PID:g603266"  
/translation="MRSFKAHKKSTFDESPKHSNFTSGNTNNSRSSDDSLDPLP  
STPSOMYDSIPPAKHSPGFSEFHLANKTSKLFKTSNSLNLSHLASTPTSTNOT  
TNSFVLQPPKNTGPPPLPPLPSPSSTSFSRHNSEYTKKTSKPAKDNRT  
TSLPAIKGIIHWSGDSKSHVILINDPASNPASNTSEATSKOFTPIGNELTS  
TTPSLEPAIRILNKKQKQENIDDAEDSGSKKEHHVYKALAKARNRQARISH  
DIIINGKASQMDMILAAFGSNTTINDDQSENEQTDKILDIERTVTTSTLTS  
ETTPSKPCFYQSOTLSLPIRHGDLQSSPSKYVKNDSQNETLNKKVRLSLNRKE  
EEKVSLNNSDEYSVNEKTHKANDCNDESSENGDNDHDDYDDDDDDDDDDDD  
EFSFEYAGINVTSSVYKYPEANVYIDLYEDENFDMDMNCIEDESGNEGNEI  
CGLSTRFETSLSKNVKKFNDLFLNSDDEEDGKNSNGDENSENLQKRLKNG  
KETFGNGHHDDASLGTVDNKEQFLINDNVKKPIQKINDLFDLSDDDNDKEMS  
TAIKASLTGTGSKTPKSFSDIFNVDSADSAESDGGTGNNSGLVNSDSROV  
EASYSFDEAPSIESPANAKSTRGIYSNKNRIIDGKPNYSFLKRNSSDDETEH  
SIQSLYETKSESHPNHPHISOILOTPAKIVITPSVDAQSOALAITDDGDDDDDD  
SILTRPFOLIDSSHQPHYASPOYAVLNSPPLPPPARSQSLKYHLNCLDSEVP  
RWSNLFFIDEAEEDEINQKAFDFDHYDIDDEINGIDPFNDSDESDNRRRTLKS  
PLUGGSKNVSFPFSVSSFSSTKSHFNGKLTINQAKELAPMKKILKTNKVTFFN  
SNWNTYDCNLSLRKTSQMRDSKYQNHNVQNVQNPSSVLSLPQHSQHSGLDCKNDNY



AUTHORS  
TITLE  
JOURNAL

Buck, D.  
Direct Submission  
Submitted (06-NOV-1996) Wellcome Trust Genome Campus, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquires:  
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
IMPORTANT: This sequence is unfinished and does not necessarily  
represent the correct sequence. Work on the sequence is in progress  
and

## COMMENT

the release of this data is based on the understanding that the  
sequence may change as work continues. The sequence may be  
contaminated  
with foreign sequence from E.coli, yeast, vector, phage etc. Order  
of segments is not known; 800 n's separate segments. Unfinished  
sequence: dj111j24 Contig\_ID: 00116 Length: 8287 bp Unfinished  
sequence: dj111j24 Contig\_ID: 02344 Length: 13962 bp Unfinished  
sequence: dj111j24 Contig\_ID: 00594 Length: 2065 bp Unfinished  
sequence: dj111j24 Contig\_ID: 00331 Length: 10666 bp Unfinished  
sequence: dj111j24 Contig\_ID: 00023 Length: 794 bp Unfinished  
sequence: dj111j24 Contig\_ID: 00704 Length: 5939 bp Unfinished  
sequence: dj111j24 Contig\_ID: 01917 Length: 4193 bp Unfinished  
sequence: dj111j24 Contig\_ID: 01504 Length: 4636 bp Unfinished  
sequence: dj111j24 Contig\_ID: 01915 Length: 4255 bp Unfinished  
sequence: dj111j24 Contig\_ID: 02681 Length: 5936 bp Unfinished  
sequence: dj111j24 Contig\_ID: 02147 Length: 3588 bp Unfinished  
sequence: dj111j24 Contig\_ID: 02187 Length: 6518 bp Unfinished  
sequence: dj111j24 Contig\_ID: 02645 Length: 11173 bp Unfinished  
sequence: dj111j24 Contig\_ID: 01817 Length: 24371 bp Unfinished  
sequence: dj111j24 Contig\_ID: 02255 Length: 5348 bp Unfinished  
sequence: dj111j24 Contig\_ID: 02053 Length: 21019 bp.

\*\*\* WARNING: Phase 1 High Throughout Genome Sequence \*\*\*  
\*\*\*

\* This sequence is unfinished. When sequencing is complete,  
\* the sequence data presented in this record will be replaced  
\* by a single finished sequence with the same accession number.

## FEATURES

source

1..145750  
/organism="Homo sapiens"  
/clone="111j24"  
/chromosome="22"

BASE COUNT 36917 a 31170 c 31921 g 33145 t 12597 others  
ORIGIN

## Query Match

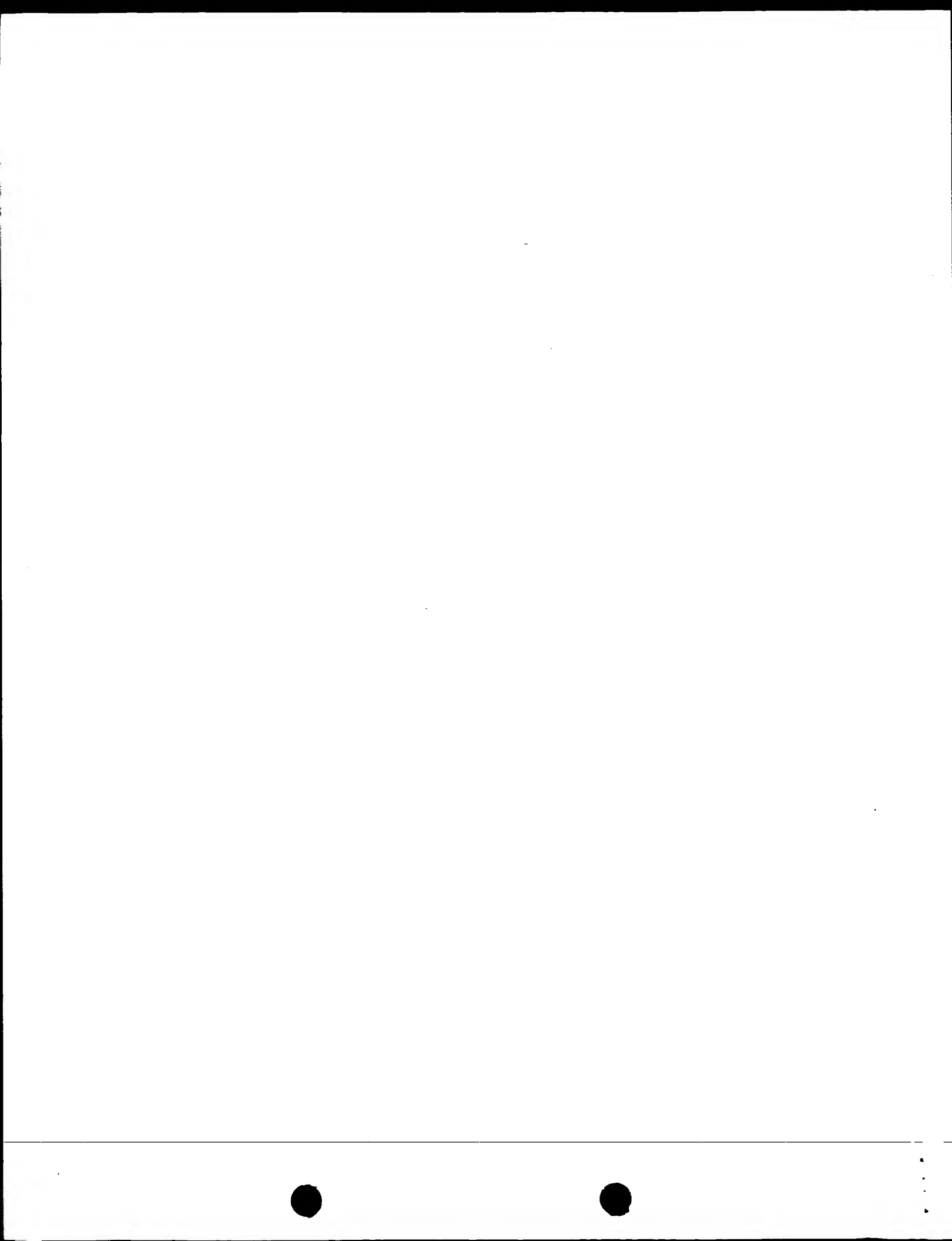
Best Local Similarity 2.1%; Score 23; DB 33; Length 145750;  
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 82692 gaacacaaacaaagaccagaat 82716

||||| ||||| ||||| ||||| |||||

771 GAACATCAAAACAAAGACCAAGAT 795

Search completed: Tue Dec 2 16:35:46 1997  
Job time : 1883 secs.





\*\*\*\*\*

WATERMAN

(TM)

\*\*\*\*\*

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by Intelligent, Inc.

MPsrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 2 17:04:56 1997; MasPar time 59.59 Seconds  
814.423 Million cell updates/sec

File: >US-08-915-004-12  
Description: (1-465) from US08915004.seq  
Perfect Score: 465  
N.A. Sequence: 1 ATGAACAAGTTGCTGTGCTG.....AGATAGTTGTGACAGTTTAG 465  
Comp: TACTTGTTCACGACGAC.....TCTATCAACACTGTCATTC

Scoring table: TABLE default  
Gap 6

Nmatch STD : Dbase 0; Query 0  
Searched: 142080 seqs, 52183452 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: n-geneseq28  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29

Statistics: Mean 8.278; Variance 4.498; scale 1.841

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	465	100.0	465	28	Osteoclastogenesis in	0.00e+00
2	437	94.0	10190	27	Fragment of human OCIF	0.00e+00
3	400	86.0	1173	28	Human tumour necrosis	2.10e+285
4	399	85.8	438	28	Osteoclastogenesis in	1.28e-284
5	398	85.6	432	27	Mutated OCIF, OCIF-DC	7.76e-284
6	398	85.6	564	27	Mutated OCIF, OCIF-CD	7.76e-284
7	398	85.6	594	27	Mutated OCIF, OCIF-CD	7.76e-284
8	398	85.6	819	27	Mutated OCIF, OCIF-CD	7.76e-284
9	398	85.6	966	27	Mutated OCIF, OCIF-CD	7.76e-284
10	398	85.6	981	27	Mutated OCIF, OCIF-CD	7.76e-284
11	398	85.6	984	27	Mutated OCIF, OCIF-CD	7.76e-284
12	398	85.6	1056	27	Mutated OCIF, OCIF-CD	7.76e-284
13	398	85.6	1080	27	Mutated OCIF, OCIF-DC	7.76e-284
14	398	85.6	1182	27	Mutated OCIF, OCIF-DC	7.76e-284
15	398	85.6	1200	27	Mutated OCIF, OCIF-CL	7.76e-284

16	398	85.6	1206	28	T36685	Osteoclastogenesis in	7.76e-284
17	398	85.6	1206	27	T33164	Mutated OCIF, OCIF-C2	7.76e-284
18	396	85.2	1206	28	T33163	Mutated OCIF, OCIF-C1	2.87e-282
19	396	85.2	1206	27	T33163	Mutated OCIF, OCIF-C2	2.87e-282
20	396	85.2	1206	27	T33165	Mutated OCIF, OCIF-C2	2.87e-282
21	396	85.2	1206	28	T33162	Mutated OCIF, OCIF-C2	2.87e-282
22	316	68.0	321	27	T33177	Mutated OCIF, OCIF-CC	1.03e-219
23	312	67.1	1080	27	T33168	Mutated OCIF, OCIF-DC	1.35e-216
24	242	52.0	255	27	T33181	Mutated OCIF, OCIF-CP	3.30e-162
25	213	45.8	1083	27	T33166	Mutated OCIF, OCIF-DC	8.17e-140
26	185	39.8	1080	27	T33167	Mutated OCIF, OCIF-DC	2.66e-118
27	49	10.5	1047	2	Q10572	Human Natriuretic Pep	9.50e-18
28	48	10.3	1047	2	Q10572	Human Natriuretic Pep	4.48e-17
29	42	9.0	204	1	N81164	Base substituted E.co	4.30e-13
30	40	8.6	91	9	Q51746	Oligonucleotide probe	8.61e-12
31	37	8.0	91	9	Q51746	Oligonucleotide probe	7.23e-10
32	37	8.0	204	1	N81164	Base substituted E.co	7.23e-10
33	36	7.7	114	12	Q70470	Generic DNA sequence	3.11e-09
34	36	7.7	114	12	Q70469	Generic DNA sequence	3.11e-09
35	36	7.7	114	12	Q70465	Generic DNA sequence	3.11e-09
36	35	7.5	114	12	Q70467	Generic DNA sequence	1.32e-08
37	34	7.3	114	12	Q70467	Generic DNA sequence	5.55e-08
38	33	7.1	114	12	Q70468	Generic DNA sequence	2.31e-07
39	33	7.1	114	12	Q70466	Generic DNA sequence	2.31e-07
40	33	7.1	114	12	Q70468	Generic DNA sequence	2.31e-07
41	32	6.9	114	12	Q70473	Generic DNA sequence	9.45e-07
42	32	6.9	114	12	Q70469	Generic DNA sequence	9.45e-07
43	32	6.9	114	12	Q70465	Generic DNA sequence	9.45e-07
44	31	6.7	114	12	Q70472	Generic DNA sequence	3.83e-06
45	30	6.5	114	12	Q70471	Generic DNA sequence	1.53e-05

ALIGNMENTS

RESULT 1  
ID T36688 standard; DNA; 465 BP.  
AC T36688;  
DT 22-APR-1997 (first entry)  
DE Osteoclastogenesis inhibitory factor 4 coding sequence.  
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; Bone resorption;  
OS Homo sapiens; ss.  
FH Key Location/Qualifiers  
FT sig\_peptide 1..63  
FT \*tag= a  
FT mat\_peptide 64..462  
FT \*tag= b  
FT \*tag= c  
FT \*tag= d  
FT \*tag= e  
FT \*tag= f  
FT \*tag= g  
FT \*tag= h  
FT \*tag= i  
FT \*tag= j  
FT \*tag= k  
FT \*tag= l  
FT \*tag= m  
FT \*tag= n  
FT \*tag= o  
FT \*tag= p  
FT \*tag= q  
FT \*tag= r  
FT \*tag= s  
FT \*tag= t  
FT \*tag= u  
FT \*tag= v  
FT \*tag= w  
FT \*tag= x  
FT \*tag= y  
FT \*tag= z  
FT \*tag= AA  
FT \*tag= AB  
FT \*tag= AC  
FT \*tag= AD  
FT \*tag= AE  
FT \*tag= AF  
FT \*tag= AG  
FT \*tag= AH  
FT \*tag= AI  
FT \*tag= AJ  
FT \*tag= AK  
FT \*tag= AL  
FT \*tag= AM  
FT \*tag= AN  
FT \*tag= AO  
FT \*tag= AP  
FT \*tag= AQ  
FT \*tag= AR  
FT \*tag= AS  
FT \*tag= AT  
FT \*tag= AU  
FT \*tag= AV  
FT \*tag= AW  
FT \*tag= AX  
FT \*tag= AY  
FT \*tag= AZ  
FT \*tag= BA  
FT \*tag= BB  
FT \*tag= BC  
FT \*tag= BD  
FT \*tag= BE  
FT \*tag= BF  
FT \*tag= BG  
FT \*tag= BH  
FT \*tag= BI  
FT \*tag= BJ  
FT \*tag= BK  
FT \*tag= BL  
FT \*tag= BM  
FT \*tag= BN  
FT \*tag= BO  
FT \*tag= BP  
FT \*tag= BQ  
FT \*tag= BR  
FT \*tag= BS  
FT \*tag= BT  
FT \*tag= BU  
FT \*tag= BV  
FT \*tag= BW  
FT \*tag= BX  
FT \*tag= BY  
FT \*tag= BZ  
FT \*tag= CA  
FT \*tag= CB  
FT \*tag= CC  
FT \*tag= CD  
FT \*tag= CE  
FT \*tag= CF  
FT \*tag= CG  
FT \*tag= CH  
FT \*tag= CI  
FT \*tag= CJ  
FT \*tag= CK  
FT \*tag= CL  
FT \*tag= CM  
FT \*tag= CN  
FT \*tag= CO  
FT \*tag= CP  
FT \*tag= CQ  
FT \*tag= CR  
FT \*tag= CS  
FT \*tag= CT  
FT \*tag= CU  
FT \*tag= CV  
FT \*tag= CW  
FT \*tag= CX  
FT \*tag= CY  
FT \*tag= CZ  
FT \*tag= DA  
FT \*tag= DB  
FT \*tag= DC  
FT \*tag= DD  
FT \*tag= DE  
FT \*tag= DF  
FT \*tag= DG  
FT \*tag= DH  
FT \*tag= DI  
FT \*tag= DJ  
FT \*tag= DK  
FT \*tag= DL  
FT \*tag= DM  
FT \*tag= DN  
FT \*tag= DO  
FT \*tag= DP  
FT \*tag= DQ  
FT \*tag= DR  
FT \*tag= DS  
FT \*tag= DT  
FT \*tag= DU  
FT \*tag= DV  
FT \*tag= DW  
FT \*tag= DX  
FT \*tag= DY  
FT \*tag= DZ  
FT \*tag= EA  
FT \*tag= EB  
FT \*tag= EC  
FT \*tag= ED  
FT \*tag= EE  
FT \*tag= EF  
FT \*tag= EG  
FT \*tag= EH  
FT \*tag= EI  
FT \*tag= EJ  
FT \*tag= EK  
FT \*tag= EL  
FT \*tag= EM  
FT \*tag= EN  
FT \*tag= EO  
FT \*tag= EP  
FT \*tag= EQ  
FT \*tag= ER  
FT \*tag= ES  
FT \*tag= ET  
FT \*tag= EU  
FT \*tag= EV  
FT \*tag= EW  
FT \*tag= EX  
FT \*tag= EY  
FT \*tag= EZ  
FT \*tag= FA  
FT \*tag= FB  
FT \*tag= FC  
FT \*tag= FD  
FT \*tag= FE  
FT \*tag= FF  
FT \*tag= FG  
FT \*tag= FH  
FT \*tag= FI  
FT \*tag= FJ  
FT \*tag= FK  
FT \*tag= FL  
FT \*tag= FM  
FT \*tag= FN  
FT \*tag= FO  
FT \*tag= FP  
FT \*tag= FQ  
FT \*tag= FR  
FT \*tag= FS  
FT \*tag= FT  
FT \*tag= FU  
FT \*tag= FV  
FT \*tag= FW  
FT \*tag= FX  
FT \*tag= FY  
FT \*tag= FZ  
FT \*tag= GA  
FT \*tag= GB  
FT \*tag= GC  
FT \*tag= GD  
FT \*tag= GE  
FT \*tag= GF  
FT \*tag= GG  
FT \*tag= GH  
FT \*tag= GI  
FT \*tag= GJ  
FT \*tag= GK  
FT \*tag= GL  
FT \*tag= GM  
FT \*tag= GN  
FT \*tag= GO  
FT \*tag= GP  
FT \*tag= GQ  
FT \*tag= GR  
FT \*tag= GS  
FT \*tag= GT  
FT \*tag= GU  
FT \*tag= GV  
FT \*tag= GW  
FT \*tag= GX  
FT \*tag= GY  
FT \*tag= GZ  
FT \*tag= HA  
FT \*tag= HB  
FT \*tag= HC  
FT \*tag= HD  
FT \*tag= HE  
FT \*tag= HF  
FT \*tag= HG  
FT \*tag= HH  
FT \*tag= HI  
FT \*tag= HJ  
FT \*tag= HK  
FT \*tag= HL  
FT \*tag= HM  
FT \*tag= HN  
FT \*tag= HO  
FT \*tag= HP  
FT \*tag= HQ  
FT \*tag= HR  
FT \*tag= HS  
FT \*tag= HT  
FT \*tag= HU  
FT \*tag= HV  
FT \*tag= HW  
FT \*tag= HX  
FT \*tag= HY  
FT \*tag= HZ  
FT \*tag= IA  
FT \*tag= IB  
FT \*tag= IC  
FT \*tag= ID  
FT \*tag= IE  
FT \*tag= IF  
FT \*tag= IG  
FT \*tag= IH  
FT \*tag= II  
FT \*tag= IJ  
FT \*tag= IK  
FT \*tag= IL  
FT \*tag= IM  
FT \*tag= IN  
FT \*tag= IO  
FT \*tag= IP  
FT \*tag= IQ  
FT \*tag= IR  
FT \*tag= IS  
FT \*tag= IT  
FT \*tag= IU  
FT \*tag= IV  
FT \*tag= IW  
FT \*tag= IX  
FT \*tag= IY  
FT \*tag= IZ  
FT \*tag= JA  
FT \*tag= JB  
FT \*tag= JC  
FT \*tag= JD  
FT \*tag= JE  
FT \*tag= JF  
FT \*tag= JG  
FT \*tag= JH  
FT \*tag= JI  
FT \*tag= JJ  
FT \*tag= JK  
FT \*tag= JL  
FT \*tag= JM  
FT \*tag= JN  
FT \*tag= JO  
FT \*tag= JP  
FT \*tag= JQ  
FT \*tag= JR  
FT \*tag= JS  
FT \*tag= JT  
FT \*tag= JU  
FT \*tag= JV  
FT \*tag= JW  
FT \*tag= JX  
FT \*tag= JY  
FT \*tag= JZ  
FT \*tag= KA  
FT \*tag= KB  
FT \*tag= KC  
FT \*tag= KD  
FT \*tag= KE  
FT \*tag= KF  
FT \*tag= KG  
FT \*tag= KH  
FT \*tag= KI  
FT \*tag= KJ  
FT \*tag= KK  
FT \*tag= KL  
FT \*tag= KM  
FT \*tag= KN  
FT \*tag= KO  
FT \*tag= KP  
FT \*tag= KQ  
FT \*tag= KR  
FT \*tag= KS  
FT \*tag= KT  
FT \*tag= KU  
FT \*tag= KV  
FT \*tag= KW  
FT \*tag= KX  
FT \*tag= KY  
FT \*tag= KZ  
FT \*tag= LA  
FT \*tag= LB  
FT \*tag= LC  
FT \*tag= LD  
FT \*tag= LE  
FT \*tag= LF  
FT \*tag= LG  
FT \*tag= LH  
FT \*tag= LI  
FT \*tag= LJ  
FT \*tag= LK  
FT \*tag= LL  
FT \*tag= LM  
FT \*tag= LN  
FT \*tag= LO  
FT \*tag= LP  
FT \*tag= LQ  
FT \*tag= LR  
FT \*tag= LS  
FT \*tag= LT  
FT \*tag= LU  
FT \*tag= LV  
FT \*tag= LW  
FT \*tag= LX  
FT \*tag= LY  
FT \*tag= LZ  
FT \*tag= MA  
FT \*tag= MB  
FT \*tag= MC  
FT \*tag= MD  
FT \*tag= ME  
FT \*tag= MF  
FT \*tag= MG  
FT \*tag= MH  
FT \*tag= MI  
FT \*tag= MJ  
FT \*tag= MK  
FT \*tag= ML  
FT \*tag= MN  
FT \*tag= MO  
FT \*tag= MP  
FT \*tag= MQ  
FT \*tag= MR  
FT \*tag= MS  
FT \*tag= MT  
FT \*tag= MU  
FT \*tag= MV  
FT \*tag= MW  
FT \*tag= MX  
FT \*tag= MY  
FT \*tag= MZ  
FT \*tag= NA  
FT \*tag= NB  
FT \*tag= NC  
FT \*tag= ND  
FT \*tag= NE  
FT \*tag= NF  
FT \*tag= NG  
FT \*tag= NH  
FT \*tag= NI  
FT \*tag= NJ  
FT \*tag= NK  
FT \*tag= NL  
FT \*tag= NM  
FT \*tag= NO  
FT \*tag= NP  
FT \*tag= NQ  
FT \*tag= NR  
FT \*tag= NS  
FT \*tag= NT  
FT \*tag= NU  
FT \*tag= NV  
FT \*tag= NW  
FT \*tag= NX  
FT \*tag= NY  
FT \*tag= NZ  
FT \*tag= OA  
FT \*tag= OB  
FT \*tag= OC  
FT \*tag= OD  
FT \*tag= OE  
FT \*tag= OF  
FT \*tag= OG  
FT \*tag= OH  
FT \*tag= OI  
FT \*tag= OJ  
FT \*tag= OK  
FT \*tag= OL  
FT \*tag= OM  
FT \*tag= ON  
FT \*tag= OO  
FT \*tag= OP  
FT \*tag= OQ  
FT \*tag= OR  
FT \*tag= OS  
FT \*tag= OT  
FT \*tag= OU  
FT \*tag= OV  
FT \*tag= OW  
FT \*tag= OX  
FT \*tag= OY  
FT \*tag= OZ  
FT \*tag= PA  
FT \*tag= PB  
FT \*tag= PC  
FT \*tag= PD  
FT \*tag= PE  
FT \*tag= PF  
FT \*tag= PG  
FT \*tag= PH  
FT \*tag= PI  
FT \*tag= PJ  
FT \*tag= PK  
FT \*tag= PL  
FT \*tag= PM  
FT \*tag= PN  
FT \*tag= PO  
FT \*tag= PP  
FT \*tag= PQ  
FT \*tag= PR  
FT \*tag= PS  
FT \*tag= PT  
FT \*tag= PU  
FT \*tag= PV  
FT \*tag= PW  
FT \*tag= PX  
FT \*tag= PY  
FT \*tag= PZ  
FT \*tag= QA  
FT \*tag= QB  
FT \*tag= QC  
FT \*tag= QD  
FT \*tag= QE  
FT \*tag= QF  
FT \*tag= QG  
FT \*tag= QH  
FT \*tag= QI  
FT \*tag= QJ  
FT \*tag= QK  
FT \*tag= QL  
FT \*tag= QM  
FT \*tag= QN  
FT \*tag= QO  
FT \*tag= QP  
FT \*tag= QQ  
FT \*tag= QR  
FT \*tag= QS  
FT \*tag= QT  
FT \*tag= QU  
FT \*tag= QV  
FT \*tag= QW  
FT \*tag= QX  
FT \*tag= QY  
FT \*tag= QZ  
FT \*tag= RA  
FT \*tag= RB  
FT \*tag= RC  
FT \*tag= RD  
FT \*tag= RE  
FT \*tag= RF  
FT \*tag= RG  
FT \*tag= RH  
FT \*tag= RI  
FT \*tag= RJ  
FT \*tag= RK  
FT \*tag= RL  
FT \*tag= RM  
FT \*tag= RN  
FT \*tag= RO  
FT \*tag= RP  
FT \*tag= RQ  
FT \*tag= RR  
FT \*tag= RS  
FT \*tag= RT  
FT \*tag= RU  
FT \*tag= RV  
FT \*tag= RW  
FT \*tag= RX  
FT \*tag= RY  
FT \*tag= RZ  
FT \*tag= SA  
FT \*tag= SB  
FT \*tag= SC  
FT \*tag= SD  
FT \*tag= SE  
FT \*tag= SF  
FT \*tag= SG  
FT \*tag= SH  
FT \*tag= SI  
FT \*tag= SJ  
FT \*tag= SK  
FT \*tag= SL  
FT \*tag= SM  
FT \*tag= SN  
FT \*tag= SO  
FT \*tag= SP  
FT \*tag= SQ  
FT \*tag= SR  
FT \*tag= SS  
FT \*tag= ST  
FT \*tag= SU  
FT \*tag= SV  
FT \*tag= SW  
FT \*tag= SX  
FT \*tag= SY  
FT \*tag= SZ  
FT \*tag= TA  
FT \*tag= TB  
FT \*tag= TC  
FT \*tag= TD  
FT \*tag= TE  
FT \*tag= TF  
FT \*tag= TG  
FT \*tag= TH  
FT \*tag= TI  
FT \*tag= TJ  
FT \*tag= TK  
FT \*tag= TL  
FT \*tag= TM  
FT \*tag= TN  
FT \*tag= TO  
FT \*tag= TP  
FT \*tag= TQ  
FT \*tag= TR  
FT \*tag= TS  
FT \*tag= TT  
FT \*tag= TU  
FT \*tag= TV  
FT \*tag= TW  
FT \*tag= TX  
FT \*tag= TY  
FT \*tag= TZ  
FT \*tag= UA  
FT \*tag= UB  
FT \*tag= UC  
FT \*tag= UD  
FT \*tag= UE  
FT \*tag= UF  
FT \*tag= UG  
FT \*tag= UH  
FT \*tag= UI  
FT \*tag= UJ  
FT \*tag= UK  
FT \*tag= UL  
FT \*tag= UM  
FT \*tag= UN  
FT \*tag= UO  
FT \*tag= UP  
FT \*tag= UQ  
FT \*tag= UR  
FT \*tag= US  
FT \*tag= UT  
FT \*tag= UU  
FT \*tag= UV  
FT \*tag= UW  
FT \*tag= UX  
FT \*tag= UY  
FT \*tag= UZ  
FT \*tag= VA  
FT \*tag= VB  
FT \*tag= VC  
FT \*tag= VD  
FT \*tag= VE  
FT \*tag= VF  
FT \*tag= VG  
FT \*tag= VH  
FT \*tag= VI  
FT \*tag= VJ  
FT \*tag= VK  
FT \*tag= VL  
FT \*tag= VM  
FT \*tag= VN  
FT \*tag= VO  
FT \*tag= VP  
FT \*tag= VQ  
FT \*tag= VR  
FT \*tag= VS  
FT \*tag= VT  
FT \*tag= VU  
FT \*tag= VV  
FT \*tag= VW  
FT \*tag= VX  
FT \*tag= VY  
FT \*tag= VZ  
FT \*tag= WA  
FT \*tag= WB  
FT \*tag= WC  
FT \*tag= WD  
FT \*tag= WE  
FT \*tag= WF  
FT \*tag= WG  
FT \*tag= WH  
FT \*tag= WI  
FT \*tag= WJ  
FT \*tag= WK  
FT \*tag= WL  
FT \*tag= WM  
FT \*tag= WN  
FT \*tag= WO  
FT \*tag= WP  
FT \*tag= WQ  
FT \*tag= WR  
FT \*tag= WS  
FT \*tag= WT  
FT \*tag= WU  
FT \*tag= WV  
FT \*tag= WW  
FT \*tag= WX  
FT \*tag= WY  
FT \*tag= WZ  
FT \*tag= XA  
FT \*tag= XB  
FT \*tag= XC  
FT \*tag= XD  
FT \*tag= XE  
FT \*tag= XF  
FT \*tag= XG  
FT \*tag= XH  
FT \*tag= XI  
FT \*tag= XJ  
FT \*tag= XK  
FT \*tag= XL  
FT \*tag= XM  
FT \*tag= XN  
FT \*tag= XO  
FT \*tag= XP  
FT \*tag= XQ  
FT \*tag= XR  
FT \*tag= XS  
FT \*tag= XT  
FT \*tag= XU  
FT \*tag= XV  
FT \*tag= XW  
FT \*tag= XX  
FT \*tag= XY  
FT \*tag= XZ  
FT \*tag= YA  
FT \*tag= YB  
FT \*tag= YC  
FT \*tag= YD  
FT \*tag= YE  
FT \*tag= YF  
FT \*tag= YG  
FT \*tag= YH  
FT \*tag= YI  
FT \*tag= YJ  
FT \*tag= YK  
FT \*tag= YL  
FT \*tag= YM  
FT \*tag= YN  
FT \*tag= YO  
FT \*tag= YP  
FT \*tag= YQ  
FT \*tag= YR  
FT \*tag= YS  
FT \*tag= YT  
FT \*tag= YU  
FT \*tag= YV  
FT \*tag= YW  
FT \*tag= YX  
FT \*tag= YY  
FT \*tag= YZ  
FT \*tag= ZA  
FT \*tag= ZB  
FT \*tag= ZC  
FT \*tag= ZD  
FT \*tag= ZE  
FT \*tag= ZF  
FT \*tag= ZG  
FT \*tag= ZH  
FT \*tag= ZI  
FT \*tag= ZJ  
FT \*tag= ZK  
FT \*tag= ZL  
FT \*tag= ZM  
FT \*tag= ZN  
FT \*tag= ZO  
FT \*tag= ZP  
FT \*tag= ZQ  
FT \*tag= ZR  
FT \*tag= ZS  
FT \*tag= ZT  
FT \*tag= ZU  
FT \*tag= ZV  
FT \*tag= ZW  
FT \*tag= ZX  
FT \*tag= ZY  
FT \*tag= ZZ

Query Match 100.0%; Score 465; DB 28; Length 465;  
Best Local Similarity 100.0%; Pred. No. 0.00e+00;























\*\*\*\*\*

\*\*\*\*\*  
M O S E R L E  
\*\*\*\*\*

(TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by Intelligenetics, Inc.

MPSrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm  
Run on: Tue Dec 2 17:06:23 1997; MasPar time 292.48 Seconds  
816.105 Million cell updates/sec  
Circular output not generated.

>US-08-915-004-12  
Description: (1-465) from US08915004.seq  
Perfect Score: 465  
N.A. Sequence: 1 ATGAACAAGTTGCTGTGCTG.....AGATAGTTGTGACAGTTTAG 465  
Comp: TACTTGTTCACGACACGAC.....TCTATCAACACTGTCATTC

Scoring table: TABLE default  
Gap 6  
Nmatch STD : Dbase 0; Query 0  
Searched: 707517 seqs, 256659390 bases x 2  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: EST-STS  
1:EST1 2:EST2 3:EST3 4:EST4 5:EST5 6:EST6 7:EST7 8:EST8  
9:EST9 10:EST10 11:EST11 12:EST12 13:EST13 14:EST14  
15:EST15 16:EST16 17:EST17 18:EST18 19:EST19 20:EST20  
21:EST21 22:EST22 23:EST23 24:EST24 25:EST25 26:EST26  
27:EST27 28:EST28 29:EST29 30:EST30 31:EST31 32:EST32  
33:EST33 34:EST34 35:EST35 36:EST36 37:EST37 38:EST38  
39:EST39 40:EST40 41:EST41 42:EST42 43:EST43 44:EST44  
45:EST45 46:EST46 47:EST47 48:EST48 49:EST49 50:EST50  
51:EST51 52:EST52 53:EST53 54:EST54 55:EST55 56:EST56  
57:EST57 58:EST58 59:EST59 60:EST60 61:EST61 62:EST62  
63:EST63 64:EST64 65:EST65 66:EST66 67:EST67 68:EST68  
69:EST69 70:EST70 71:EST71 72:EST72 73:EST73 74:EST74  
75:EST75 76:EST76 77:EST77 78:EST78 79:EST79 80:EST80  
81:EST81 82:EST82 83:EST83 84:EST84 85:EST85 86:EST86  
87:EST87 88:EST88 89:EST89 90:EST90 91:EST91 92:EST92  
93:EST93 94:EST94 95:EST95 96:EST96 97:EST97 98:EST98  
99:EST99  
EST-STS-TWO  
100:EST100 101:EST101 102:EST102 103:EST103 104:EST104  
105:EST105 106:EST106 107:EST107 108:EST108 109:EST109  
110:EST110 111:EST111 112:EST112 113:EST113 114:EST114  
115:EST115 116:EST116 117:EST117 118:EST118 119:EST119  
120:EST120 121:EST121 122:EST122 123:EST123 124:EST124  
125:EST125 126:EST126 127:EST127 128:EST128 129:EST129  
130:EST130 131:EST131 132:EST132 133:EST133 134:EST134  
135:EST135 136:EST136 137:EST137 138:EST138 139:EST139  
140:EST140 141:EST141 142:EST142 143:EST143 144:EST144  
145:EST145 146:EST146 147:EST147 148:EST148 149:EST149  
150:EST150 151:EST151 152:EST152 153:EST153 154:EST154  
155:EST155 156:EST156 157:EST157 158:EST158 159:EST159  
160:EST160 161:EST161 162:EST162 163:EST163 164:EST164  
165:EST165 166:EST166 167:EST167 168:EST168 169:EST169  
170:EST170 171:EST171 172:EST172 173:EST173 174:EST174

175:EST175 176:EST176 177:EST177 178:EST178 179:EST179  
180:EST180 181:EST181 182:EST182 183:EST183 184:EST184  
185:EST185 186:EST186 187:EST187 188:EST188 189:EST189  
190:EST190 191:EST191 192:EST192 193:EST193 194:EST194  
195:EST195 196:EST196 197:EST197 198:EST198

Statistics: Mean 10.073; Variance 1.752; scale 5.751

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description	Pred. No.
1	23	4.9	344 61	H14106	ym62a05.r1 Homo sapie	1.72e-07
2	22	4.7	300 183	AA100384	zn46h08.r1 Strataene	4.46e-06
3	22	4.7	453 111	N21157	yx47d01.s1 Homo sapie	4.46e-06
4	22	4.7	472 57	T42477	5740 Arabidopsis thal	4.46e-06
5	22	4.7	433 55	RICS15559A	Rice cDNA, partial se	4.46e-06
6	22	4.7	796 117	W29026	55c10 Human retina cd	4.46e-06
7	21	4.5	300 48	HUM213B09B	Human aorta cDNA 5'-e	1.03e-04
8	21	4.5	325 19	T54964	yb42d03.r1 Homo sapie	1.03e-04
9	21	4.5	371 99	N61165	TgESTzy27b03.r1 Toxop	1.03e-04
10	21	4.5	376 5	T71088	Y50404.r1 Homo sapie	1.03e-04
11	21	4.5	383 47	H45707	Yp23h05.s1 Homo sapie	1.03e-04
12	21	4.5	393 5	T71079	Y50C04.r1 Homo sapie	1.03e-04
13	21	4.5	447 156	AA021559	ze69b06.s1 Soares ret	1.03e-04
14	21	4.5	447 184	AA011021	ze34c01.s1 Soares ret	1.03e-04
15	21	4.5	491 183	AA098806	zn44d05.s1 Strataene	1.03e-04
16	20	4.3	218 75	HUM028E02B	Human fetal brain CDN	2.11e-03
17	20	4.3	228 50	R83564	Yql2e06.r1 Homo sapie	2.11e-03
18	20	4.3	233 55	RICS11807A	Rice cDNA, partial se	2.11e-03
19	20	4.3	271 4	T69650	Yc44c09.r1 Homo sapie	2.11e-03
20	20	4.3	292 109	HSIMBB115	H.sapiens partial mRN	2.11e-03
21	20	4.3	301 140	C01787	HUMGS0003743, Human G	2.11e-03
22	20	4.3	323 5	T71171	YC52c03.r1 Homo sapie	2.11e-03
23	20	4.3	349 5	T72764	YC51a10.r1 Homo sapie	2.11e-03
24	20	4.3	369 71	H83210	Yq47d03.r1 Homo sapie	2.11e-03
25	20	4.3	371 173	W72637	zd70e07.s1 Soares fet	2.11e-03
26	20	4.3	374 78	R98014	Yq57f02.s1 Homo sapie	2.11e-03
27	20	4.3	375 146	W06794	SMEST0430 Schlitosoma	2.11e-03
28	20	4.3	377 177	AA067387	26232 Lambda-PRL2 Ara	2.11e-03
29	20	4.3	377 2	T61303	yb84f02.r1 Homo sapie	2.11e-03
30	20	4.3	379 178	AA072548	MB4SLD058r3 Brugia ma	2.11e-03
31	20	4.3	390 174	W79653	zd73h08.r1 Soares fet	2.11e-03
32	20	4.3	411 174	W79538	zd81g01.s1 Soares fet	2.11e-03
33	20	4.3	411 17	T48207	yb44a06.s1 Homo sapie	2.11e-03
34	20	4.3	412 185	AA135635	zl24d10.s1 Soares pre	2.11e-03
35	20	4.3	413 120	W51540	AS4SLC915r3 Brugia ma	2.11e-03
36	20	4.3	420 197	H94319	Yv18b09.s1 Soares fet	2.11e-03
37	20	4.3	429 53	R92022	Yp96e05.s1 Homo sapie	2.11e-03
38	20	4.3	429 75	H97994	Yw06e04.s1 Homo sapie	2.11e-03
39	20	4.3	435 68	H72171	Yr99g06.r1 Homo sapie	2.11e-03
40	20	4.3	441 165	C20472	Rice cDNA, partial se	2.11e-03
41	20	4.3	444 165	C20089	Rice cDNA, partial se	2.11e-03
42	20	4.3	453 104	N74503	za54h03.s1 Homo sapie	2.11e-03
43	20	4.3	457 74	H93069	Yw06h02.s1 Homo sapie	2.11e-03
44	20	4.3	460 193	AA166259	ms49d11.r1 Life Tech	2.11e-03
45	20	4.3	531 115	W20309	zb43g03.r1 Soares fet	2.11e-03

ALIGNMENTS

RESULT 1 H14106 344 bp mRNA EST 10-JUL-1995  
LOCUS ym62a05.r1 Homo sapiens cDNA clone 163472 5' similar to SP:532367  
DEFINITION S32367 ALPA-SNAP PROTEIN - ;  
ACCESSION H14106  
NID 9878954  
KEYWORDS EST.

## SOURCE

human clone=163472 library=Soares adult brain N2b4HB53Y  
vector=pt7T3D (Pharmacia) with a modified polylinker host=DH10B  
(ampicillin resistant) primer=M13Rpl Rsite1-Not I Rsite2-Eco RI  
55-year old male. 1st strand cDNA was primed with a Not I -  
oligo(dT) primer [5'  
TGTTCACCAATGCTGAAGTGGAGCGCCGCGCTTTTTTTTTTTT 3'],  
(Pharmacia), digested with Not I and cloned into the Not I and Eco  
RI sites of a modified pT73 vector (Pharmacia). Library went  
through one round of normalization to a Cot -53. Library  
constructed by Bento Soares and M.Patima Bonaldo. The adult brain  
RNA was provided by Dr. Donald H. Gilden. Tissue was acquired 17-18  
hours after death which occurred in consequence of a ruptured  
aortic aneurysm. RNA was prepared from a pool of tissues  
representing the following areas of the brain: frontal, parietal,  
temporal and occipital cortex from the left and right hemispheres,  
subcortical white matter, basal ganglia, thalamus, cerebellum,  
midbrain, pons and medulla.

## ORGANISM

Homo sapiens  
Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;  
Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;  
Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria;  
Eutheria; Archonta; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 344)  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and  
Wilson, R.  
The WashU-Merck EST Project  
Unpublished (1995)

## TITLE

WashU-Merck EST Project  
Unpublished (1995)

## FEATURES

source  
1..344  
/organism="Homo sapiens"  
/clone="163472"  
71 a 116 c 86 g 64 t 7 others

## COUNT

71 a 116 c 86 g 64 t 7 others

Query Match 4.9%; Score 23; DB 61; Length 344;  
Best Local Similarity 77.8%; Pred. No. 1.72e-07;  
Matches 28; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Db 117 gcacacttccctnactactnctacncaacagct 152

Qy 185 GCGCCCTGCTGACCACTACTACACACASCT 220

## RESULT

LOCUS AA100384 300 bp mRNA EST 28-OCT-1996  
DEFINITION zn46808.r1 Stratagene HeLa cell s3 937216 Homo sapiens cDNA clone  
550527 5'.

ACCESSION AA100384

NID g1646757

KEYWORDS EST.

SOURCE human.

## ORGANISM

Homo sapiens  
Eukaryotae; Mitochondrial eukaryotes; Metazoa; Chordata;  
Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 300)  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,

## REFERENCE

AUTHORS

TITLE  
JOURNAL  
COMMENT

Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevaskis, E.,  
Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.  
WashU-Merck EST Project  
Unpublished (1995)

Contact: Wilson RK

WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: est@wustl.wustl.edu

This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -28M13 rev2 from Amersham

High quality sequence stop: 258.

## FEATURES

Location/Qualifiers

source

1..300  
/organism="Homo sapiens"  
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:  
XhoI; Cloned unidirectionally. Primer: Oligo dT. HeLa S3  
epithelioid carcinoma cells grown to semi-confluency  
without induction. Average insert size: 1.5 kb; Uni-ZAP XR  
vector. -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3'  
adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'  
/clone="550527"  
/clone.lib="Stratagene HeLa cell s3 937216"  
/sex="female"  
/dev\_stage="HeLa S3 cell line"  
/lab\_host="SOLR (kanamycin resistant)"

BASE COUNT 109 a 43 c 51 g 95 t 2 others

## ORIGIN

<1..>300

## Query Match

Best Local Similarity 4.7%; Score 22; DB 183; Length 300;

Matches 37; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Db 215 gtaacttcattgcaaccagaaacctgtaatacgtgtacagtaacaaagtgt 266

Qy 81 GTACCTTCATTATGACGAAGAAGAACTCTCATCAGCTGTGTGTGACAAATGT 132

## RESULT

LOCUS N21157 453 bp mRNA EST 19-DEC-1995  
DEFINITION Yx47d01.s1 Homo sapiens cDNA clone 264865 3' similar to  
SP:TCPB\_MOUSE P80314 T-COMPLEX PROTEIN 1, BETA SUBUNIT ;  
ACCESSION N21157  
NID g1126327  
KEYWORDS EST.  
SOURCE human clone=264865 primer-m13 -40 forward library=Soares melanocyte

28BHM vector-pt7T3D (Pharmacia) with a modified polylinker  
host=DH10B (ampicillin resistant) Rsite1-Not I Rsite2-Eco RI Male.  
1st strand cDNA was primed with a Not I - oligo(dT) primer  
[5'-TGTTACCAATCTGAAGTGGAGCGCCGCGCTTTTTTTTTTTT 3'],  
double-stranded cDNA was size selected, ligated to Eco RI adapters  
(Pharmacia), digested with Not I and cloned into the Not I and Eco  
RI sites of a modified pT73 vector (Pharmacia). Library  
constructed by Bento Soares and M.Patima Bonaldo. RNA from normal  
foreskin melanocytes (FS374) was kindly provided by Dr. Anthony P.  
Albino.

## ORGANISM

Homo sapiens  
Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;  
Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;  
Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria;  
Eutheria; Archonta; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 453)

## REFERENCE

AUTHORS

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and  
Wilson, R.  
The WashU-Merck EST Project

## TITLE



Tel: 410 955 4678  
 Fax: 410 614 0827  
 Email: jeremy.nathans@gmail.bs.jhu.edu  
 Clones from this library are NOT available.  
 PCR Primers  
 FORWARD: CTTTGGAGCAAGTTCAGCTGGTGAAGT  
 BACKWARD: GAGGTGGGTATGAGTATTTCTTCAGGGTAA  
 Seq primer: GGGTAAAGCAAAAGAATT.

FEATURES  
 source  
 Location/Qualifiers  
 1..796  
 /organism="Homo sapiens"  
 /note="Organ: eye; Vector: lambda gt10; Site: 1: EcoRI;  
 Site: 2: EcoRI; The library used for sequencing was a  
 sublibrary derived from a human retina cDNA library.  
 Inserts from retina cDNA library DNA were isolated  
 randomly primed. PCR amplified, size-selected, and cloned  
 into lambda gt10. Individual plaques were arrayed and  
 used as templates for PCR amplification, and these PCR  
 products were used for sequencing."  
 /clone\_lib="Human retina cDNA randomly primed sublibrary"  
 /sex="mixed (males and females)"  
 /tissue\_type="retina"  
 /dev\_stage="adult"  
 /lab\_host="E. coli strain K802"  
 <1..>796

BASE COUNT  
 ORIGIN  
 228 a 181 c 183 g 148 t 56 others

Query Match  
 Best Local Similarity 4.7%; Score 22; DB 117; Length 796;  
 Matches 34; Conservative 0; Mismatches 12; Indels 0; Gaps 0;  
 Db 92 gcaaggaatgcagattatgcagcagtgagctggtaccacaa 137  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Qy 260 GCAAGGAGCTGCAGTACGTCAAGCAGGAGTCAATCGCACCCACAA 305

RESULT  
 LOCUS HUM213B09B 300 bp mRNA EST 27-AUG-1995  
 DEFINITION Human aorta cDNA 5'-end GEN:213B09.  
 ACCESSION D56616  
 NID g963238  
 KEYWORDS EST(expressed sequence tag); Human aorta; similar to known(May 29,1995).  
 SOURCE Homo sapiens (library: Clontech human aorta polyA+ mRNA (#6572))  
 cDNA to mRNA.  
 ORGANISM Homo sapiens  
 Eukaryotae; Mitochondrial eukaryotes; Metazoa; Chordata;  
 Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 1 (bases 1 to 300)  
 Fujiwara, T., Hirano, H., Katagiri, T., Kawai, A., Kuga, Y., Nagata, M.,  
 Okuno, S., Ozaki, K., Shimizu, F., Shimada, Y., Shinomiya, H.,  
 Takaichi, A., Takeda, S., Watanabe, T., Takahashi, E.-I., Hirai, Y.,  
 Maekawa, H., Shin, S. and Nakamura, Y.  
 Unpublished(003)  
 TITLE Submitted (30-May-1995) to DDBJ by:  
 JOURNAL Isotomu Fujiwara  
 COMMENT Otsuka GEN Research Institute  
 Otsuka Pharmaceutical Co., Ltd  
 463-10 Kagasuno Kawauchi-cho  
 Tokushima, Tokushima  
 771-01  
 Japan  
 Phone: 0886-65-2888  
 Fax : 0886-37-1035.  
 Location/Qualifiers  
 1..300  
 /organism="Homo sapiens"  
 /clone\_lib="Clontech human aorta polyA+ mRNA (#6572)"  
 120 a 49 c 32 g 85 t 14 others

FEATURES  
 source  
 Location/Qualifiers  
 1..300  
 /organism="Homo sapiens"  
 /clone\_lib="Clontech human aorta polyA+ mRNA (#6572)"  
 120 a 49 c 32 g 85 t 14 others

Query Match  
 Best Local Similarity 4.5%; Score 21; DB 48; Length 300;  
 Matches 30; Conservative 6; Mismatches 15; Indels 0; Gaps 0;

Db 131 aahthctbagaatccatcttaataaahhcaahhcaataaacgt 181  
 ||:||||: ||| | ||| |||: ||| ||| ||| ||| |||  
 Cp 453 AACATATCGACTTGTCATGATCCCTAAATTAATTTGTCGACATTCACACGT 403

RESULT  
 LOCUS T54964 325 bp mRNA EST 08-FEB-1995  
 DEFINITION YB42D03.r1 Homo sapiens cDNA clone 73829 5' similar to  
 gb:J03910\_rnal Human (HUMAN).  
 ACCESSION T54964  
 NID g656825  
 KEYWORDS EST.  
 SOURCE human clone-73829 library-Stratagene fetal spleen (#937205)  
 vector=pluscript SK- host-SOLR cells (kanamycin resistant)  
 primer-M13R1 Rsite1-EcoRI Rsite2-XhoI Pooled fetal spleens. Cloned  
 unidirectionally. Primer: Oligo dt. Average insert size: 1.0 kb;  
 Uni-ZAP XR Vector; 5' adaptor sequence: 5'-GAATTCGGCAGCAG-3'; 3'  
 adaptor sequence: 5'-CTCGAGTCTTTTCTTTTCTTTT-3'.

ORGANISM Homo sapiens  
 Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 1 (bases 1 to 325)  
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
 Holman, J., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
 Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevisan, E.,  
 Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.  
 WashU-Merck EST Project  
 Unpublished (1995)

TITLE  
 JOURNAL  
 COMMENT

Contact: Wilson RK  
 WashU-Merck EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.edu  
 Source: IMAGE Consortium, LLNL  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Putative full length read.

FEATURES  
 source  
 1..325  
 /organism="Homo sapiens"  
 /clone="73829"  
 BASE COUNT 68 a 76 c 77 g 92 t 12 others  
 ORIGIN

Query Match  
 Best Local Similarity 4.5%; Score 21; DB 19; Length 325;  
 Matches 29; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
 Db 50 gtgcaaatgcactctgcagaagagctgctgctct 86  
 ||||| ||||| ||||| ||| ||| |||||  
 Cp 299 GTGCGATTGCATCTCTCTTCGACTGACTGAGTCTCT 263

RESULT  
 LOCUS N61165 371 bp mRNA EST 02-APR-1996  
 DEFINITION T9ES1zy27b03.r1 Toxoplasma gondii cDNA clone tgy27b03.r1 5'  
 similar to SW:EF1A\_PLAFK Q00080 ELONGATION FACTOR 1-ALPHA ;  
 ACCESSION N61165  
 NID g1207316  
 KEYWORDS EST.  
 SOURCE clone=tgy27b03.r1 primer=T3 library=TgRcDNA strain-RH  
 vector=lambda ZAP host-XL1-Blue MRF; Rsite1-EcoRI Rsite2-XhoI  
 foreskin fibroblast in vitro cultures. Directional cDNA was  
 synthesized by oligo d(T) priming and cloned into EcoRI and XhoI  
 sites of the Lambda ZAP vector using the ZAP-cDNA Synthesis Kit  
 (Stratagene).

```

FEATURES
  source
    Location/Qualifiers
      1..376
        /organism="Homo sapiens"
        /clone="84103"
      78 a 105 c 88 g 102 t 3 others
      ORIGIN

Query Match 4.5%; Score 21; DB 5; Length 376;
Best Local Similarity 78.4%; Pred. No. 1.03e-04;
Matches 29; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Ddb 122 gtgcaaatgcactctgcaagaagctgctgctct 158
||||| ||||||||| || | |||| |||||
Cp 299 GTGCATTGCACCTCCTGCTTGCAGTACTGCAGCTCT 263

RESULT 11
LOCUS H45707 383 bp mrna EST 31-JUL-1995
DEFINITION YP23H05.s1 Homo sapiens cDNA clone 188313 3'.
ACCESSION H45707
VERSION 9921759
NID
KEYWORDS EST.
SOURCE
  human clone-188313 library-Soares breast 3NDBST vector-ptT3D
  (Pharmacia) with a modified polylinker host-DH10B (ampicillin
  resistant) primer-SP6 Rsite2-Not I Rsite2-Eco RI Adult human. 1st
  strand cDNA was primed with a Not I - oligo(dT) primer [5',
  TGTTACCAATCGAATGGAGCGAGCGCCGCTTTTTTTTTTTT 3'],
  double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia),
  digested with Not I and cloned into the Not I and Eco RI sites of a
  modified pTT3 vector (Pharmacia). Library went through one round
  of normalization to a Cot = 20. Library constructed by Bento Soares
  and M.Fatima Bonaldo.

ORGANISM Homo sapiens
  Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
  Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
  Sarcopterygii; Chonadata; Tetrapoda; Amniota; Mammalia; Theria;
  Eutheria; Archonta; Primates; Catarrhini; Hominoideae; Homo.
  1 (bases 1 to 383)
  1 (Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
  Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
  Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
  Trevasakis, E., Waterston, R., Williamson, A., Wohldmann, P. and
  Wilson, R.
  The WashU-Merck EST Project
  Unpublished (1995)

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 268
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

FEATURES
  source
    Location/Qualifiers
      1..383
        /organism="Homo sapiens"
        /clone="188313"
      115 a 79 c 91 g 90 t 8 others
      ORIGIN

Query Match 4.5%; Score 21; DB 47; Length 383;
Best Local Similarity 81.8%; Pred. No. 1.03e-04;
Matches 27; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 258 cctccacattaccctcattaggacgtgaacc 290
||||| ||||||| ||||| ||||| |||||
QY 73 CCTCCAAAGTACCTTCATTATGACGAAGAACC 105

```

```

RESULT 12
LOCUS   T71079      395 bp      mRNA      EST      01-MAR-1995
DEFINITION   yc50c04.t1 Homo sapiens cDNA clone 84102 5' similar to
              gb:J03910_rnal Human (HUMAN)).
ACCESSION   T71079
NID        9685600
KEYWORDS    EST.
SOURCE      human clone-84102 library-stratagene liver (#937224)
              vector-pBluescript SK host-SOLR cells (kanamycin resistant)
              primer-M13Rpl esitel-EcoRI rsite2-XhoI Cloned unidirectionally.
              Primer: Oligo dt. Hepatotomy from normal 49 year old male
              caucasian. Average insert size: 1.1 Kb; Uni-ZAP XR Vector; 5'
              adaptor sequence: 5'-GAATTCGGCAG-3'; 3' adaptor sequence:
              5'-CTGAGTGTGTTTTTTTTTT-3'.
ORGANISM    Homo sapiens
              Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
              Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE   1 (bases 1 to 395)
AUTHORS    Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
              Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
              Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevasakis,E.,
              Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.
              WashU-Merck EST Project
              Unpublished (1995)
TITLE      Contact: Wilson RK
JOURNAL    Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 304
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES    Location/Qualifiers
              1..395
              /organism="Homo sapiens"
              /clone="84102"
BASE COUNT  85 a 105 c 91 g 110 t 4 others
ORIGIN
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```

```

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 417.
FEATURES    Location/Qualifiers
              1..447
              /organism="Homo sapiens"
              /note="Organ: eye; Vector: pT7T3D (Pharmacia) with a
              modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
              strand cDNA was primed with a Not I - oligo(dT) primer [5'
              TGTTCACCAATCTGAAGTGGAGCGCGCGCTTTTCTTTTCTTTT 3'],
              double-stranded cDNA was size selected, ligated to Eco RI
              adapters (Pharmacia), digested with Not I and cloned into
              the Not I and Eco RI sites of a modified pT7T3 vector
              (Pharmacia). The retinas were obtained from a 55 year old
              Caucasian and total cellular poly(A)+ RNA was extracted 6
              hrs after their removal. The retina RNA was kindly
              provided by Roderick R. McInnes M.D. Ph.D. from the
              University of Toronto. Library constructed by Bento
              Soares and M.Fatima Bonaldo."
              /clone="364211"
              /clone_lib="Soares retina N2b4HR"
              /sex="male"
              /tissue_type="retina"
              /dev_stage="55 year old"
              /lab_host="DH10B (ampicillin resistant)"
              complement(<1..>447)
BASE COUNT  140 a 97 c 111 t
ORIGIN
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```

Contact: Wilson RK  
 WashU-Merck EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Insert Length: 603 Std Error: 0.00



Seq primer: mob.REGA+ET  
High quality sequence stop: 415.  
Location/Qualifiers  
1..447  
/organism="Homo sapiens"  
/note="Organ: eye; Vector: pT73D (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTCACCAATCGAAGTGGAGCGCGCGCTTTTCTTTTCTTTT 3'], RI double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). The retinas were obtained from a 55 year old Caucasian and total cellular poly(A)+ RNA was extracted 6 hrs after their removal. The retina RNA was kindly provided by Roderick R. McInnes M.D. Ph.D. from the University of Toronto. Library constructed by Bento Soares and M.Fallma Bonaldo."

#### FEATURES

source

BASE COUNT 137 a 97 c 99 g 112 t 2 others  
ORIGIN

Query Match 4.5%; Score 21; DB 184; Length 447;  
Best Local Similarity 81.8%; Pred. No. 1.03e-04;  
Matches 27; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 256 cctccacattaccgtcattagcgcgtgaacc 288  
||||| | ||| ||||| ||||| |||||  
QY 73 CCTCCAAAGTACCTCATTATGACGAGAAACC 105

RESULT 15  
LOCUS AA098806 491 bp mRNA EST 28-OCT-1996  
DEFINITION zn44d05.s1 stragene HeLa cell s3 937216 Homo sapiens cDNA clone 550281 3' similar to SW:NI2M\_BOVIN Q02369 NADH-UBIQUINONE OXIDOREDUCTASE B22 SUBUNIT ;

ACCESSION AA098806  
NID 91644777  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 491)  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevas, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.  
WashU-Merck EST Project  
Unpublished (1995)

Contact: Wilson RK  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Possible reversed clone: similarity on wrong strand  
Seq primer: -40M13 fwd. from Amersham  
High quality sequence stop: 423.  
Location/Qualifiers  
1..491

#### FEATURES

source

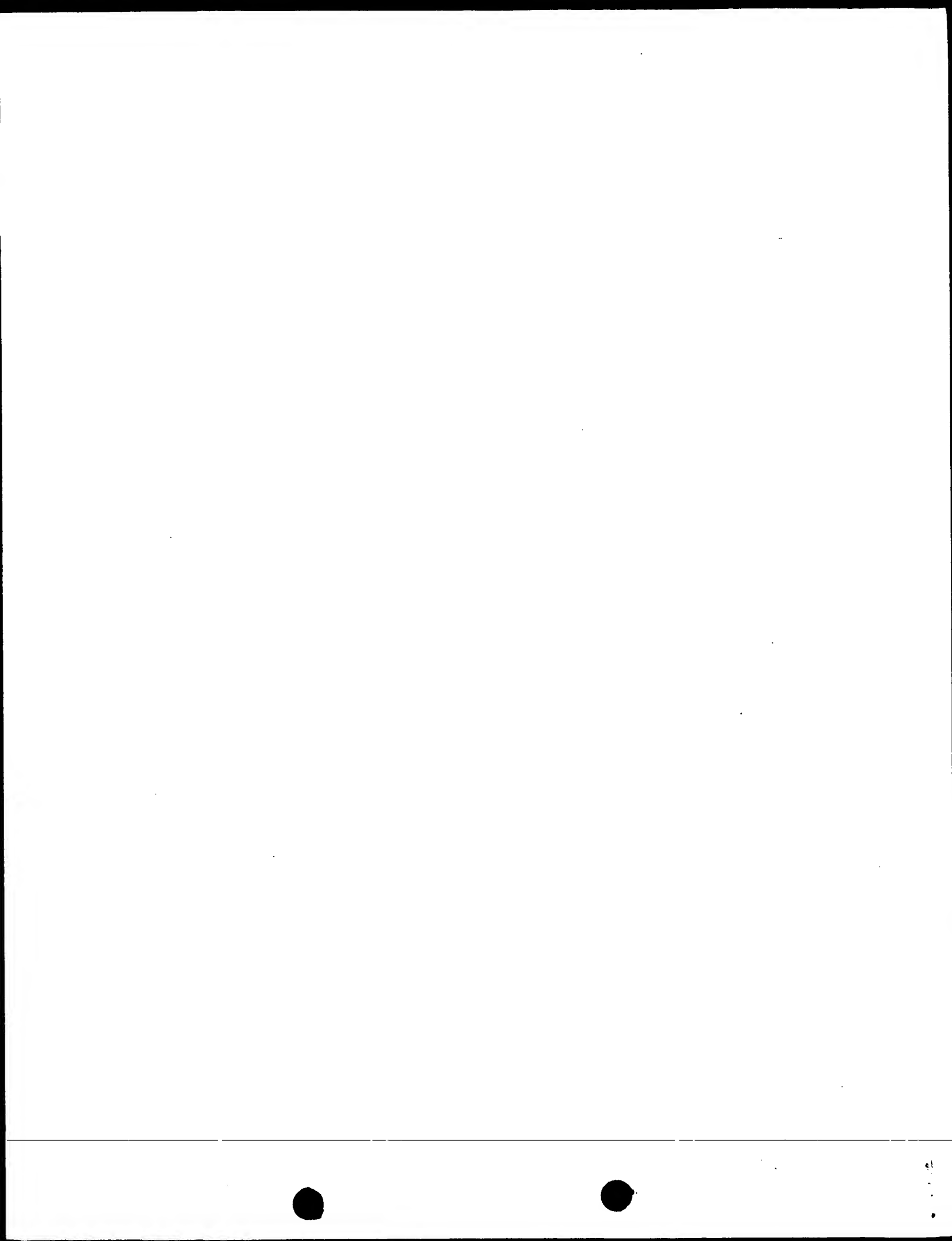
/organism="Homo sapiens"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:

XhoI; Cloned unidirectionally. Primer: Oligo dT. HeLa S3 epitheloid carcinoma cells grown to semi-confluency without induction. Average insert size: 1.5 kb; Uni-ZAP XR Vector. -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3' "  
/clone="550281"  
/clone\_lib="Stratagene HeLa cell s3 937216"  
/sex="female"  
/dev\_stage="HeLa S3 cell line"  
/lab\_host="SOLR (kanamycin resistant)"  
complement(<1..>491)  
BASE COUNT 142 a 111 c 134 g 104 t  
ORIGIN

Query Match 4.5%; Score 21; DB 183; Length 491;  
Best Local Similarity 83.9%; Pred. No. 1.03e-04;  
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 295 ggaggaaacgccacctgggtggtccttttaa 325  
||||| | ||| ||||| ||||| |||||  
Cp 77 GGAGAAACGTTTCCTGGGTGGTCCACTTAA 47

Search completed: Tue Dec 2 17:11:26 1997  
Job time : 303 secs.



\*\*\*\*\*

WAPR 1997

(TM)

\*\*\*\*\*

Release 2.1D John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by Intelligent, Inc.

MPsrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 2 17:11:48 1997; MasPar time 152.68 seconds  
Parallel output not generated. 836.954 Million cell updates/sec

Title: &gt;US-08-915-004-12

Description: (1-465) from US08915004.seq

Perfect Score: 465

N.A. Sequence: 1 ATGACAGAGTTGCTGCTG .....AGATAGTTGTCACAGTTAG 465  
Comp: TACTTGTTACAGCAGCAGC.....TCTATCAACTGTCAATC

Scoring table: TABLE default

Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 359085 seqs, 137405154 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

EST-STS-THREE  
1:EST199 2:EST200 3:EST201 4:EST202 5:EST203 6:EST204  
7:EST205 8:EST206 9:EST207 10:EST208 11:EST209 12:EST210  
13:EST211 14:EST212 15:EST213 16:EST214 17:EST215  
18:EST216 19:EST217 20:EST218 21:EST219 22:EST220  
23:EST221 24:EST222 25:EST223 26:EST224 27:EST225  
28:EST226 29:EST227 30:EST228 31:EST229 32:EST230  
33:EST231 34:EST232 35:EST233 36:EST31 37:EST32 38:EST33  
39:EST34 40:EST35 41:EST36 42:EST37 43:EST38 44:EST39 45:EST40  
46:EST41 47:EST42 48:EST43  
EST-STS-FOUR  
49:gnEST1 50:gnEST2 51:gnEST3 52:gnEST4 53:gnEST5  
54:gnEST6 55:gnEST7 56:gnEST8 57:gnEST9 58:gnEST10  
59:gnEST11 60:gnEST12 61:gnEST13 62:gnEST14 63:gnEST15  
64:gnEST16 65:gnEST17 66:gnEST18 67:gnEST19 68:gnEST20  
69:gnEST21 70:gnEST22 71:gnEST23 72:gnEST24 73:gnEST25  
74:gnEST1 75:gnEST1 76:gnEST1 77:gnEST1 78:gnEST1  
79:gnEST5 80:gnEST6 81:gnEST7 82:gnEST8 83:gnEST9  
84:gnEST10 85:gnEST11 86:gnEST12 87:gnEST13 88:gnEST14  
89:gnEST15 90:gnEST16 91:gnEST17 92:gnEST18 93:gnEST19  
94:gnEST20 95:gnEST21 96:gnEST22 97:gnEST23 98:gnEST24  
99:gnEST25 100:gnEST26 101:gnEST27 102:gnEST28  
103:gnEST29 104:gnEST30 105:gnEST31 106:gnEST32 107:gnEST33  
108:gnEST34

Database:

EST-STS-FOUR  
49:gnEST1 50:gnEST2 51:gnEST3 52:gnEST4 53:gnEST5  
54:gnEST6 55:gnEST7 56:gnEST8 57:gnEST9 58:gnEST10  
59:gnEST11 60:gnEST12 61:gnEST13 62:gnEST14 63:gnEST15  
64:gnEST16 65:gnEST17 66:gnEST18 67:gnEST19 68:gnEST20  
69:gnEST21 70:gnEST22 71:gnEST23 72:gnEST24 73:gnEST25  
74:gnEST1 75:gnEST1 76:gnEST1 77:gnEST1 78:gnEST1  
79:gnEST5 80:gnEST6 81:gnEST7 82:gnEST8 83:gnEST9  
84:gnEST10 85:gnEST11 86:gnEST12 87:gnEST13 88:gnEST14  
89:gnEST15 90:gnEST16 91:gnEST17 92:gnEST18 93:gnEST19  
94:gnEST20 95:gnEST21 96:gnEST22 97:gnEST23 98:gnEST24  
99:gnEST25 100:gnEST26 101:gnEST27 102:gnEST28  
103:gnEST29 104:gnEST30 105:gnEST31 106:gnEST32 107:gnEST33  
108:gnEST34

Statistics: Mean 10.125; Variance 1.771; scale 5.717

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result

Query

No.	Score	Match	Length	DB	ID	Description	Pred. No.
C 1	23	4.9	874	19	AA203500	zx58h07.r1 Soares fet	1.32e-07
C 2	22	4.7	472	77	AT4774	5740 Arabidopsis chal	3.31e-06
C 3	21	4.5	431	24	AA219045	zql1609.sl Stratagene	7.41e-05
C 4	20	4.3	204	106	HSC5200	Human chromosome 5 LA	1.47e-03
C 5	20	4.3	204	46	HUMC5200	Human chromosome 5 LA	1.47e-03
C 6	20	4.3	204	47	HUMC5200	Human chromosome 5 LA	1.47e-03
C 7	20	4.3	204	74	HUMC5200	Human chromosome 5 LA	1.47e-03
C 8	20	4.3	235	79	BMAA41554	MBL2S3J8A6T3 JHU96SL-B	1.47e-03
C 9	20	4.3	235	67	AA241654	MBL2S3J8A6T3 JHU96SL-B	1.47e-03
C 10	20	4.3	308	69	BMAA41559	MBL2S3J8A6T3 JHU96SL-B	1.47e-03
C 11	20	4.3	308	67	AA241559	MBL2S3J8A6T3 JHU96SL-B	1.47e-03
C 12	20	4.3	377	79	AT467387	26232 Lambda-PRL2 Ara	1.47e-03
C 13	20	4.3	383	79	BMAA41689	MBL2S3J9D7T3 JHU96SL-B	1.47e-03
C 14	20	4.3	383	67	AA241689	MBL2S3J9D7T3 JHU96SL-B	1.47e-03
C 15	20	4.3	387	79	BMAA41525	MBL2S3J9E3T3 JHU96SL-B	1.47e-03
C 16	20	4.3	387	67	AA241525	MBL2S3J9E3T3 JHU96SL-B	1.47e-03
C 17	20	4.3	393	79	BMI1391	SWAMCA485SK Brugia ma	1.47e-03
C 18	20	4.3	393	1	W15139	human STS SHGC-7175 c	1.47e-03
C 19	20	4.3	400	40	G14220	human STS SHGC-7175 c	1.47e-03
C 20	20	4.3	402	79	BMAA41471	MBL2S3J0B11T3 JHU96SL-	1.47e-03
C 21	20	4.3	402	67	AA241471	MBL2S3J0B11T3 JHU96SL-	1.47e-03
C 22	20	4.3	417	82	HS1147657	zr55g11.sl Soares Nhh	1.47e-03
C 23	20	4.3	417	63	AA228090	zr55g11.sl Soares Nhh	1.47e-03
C 24	20	4.3	421	83	HS1153832	zt30e09.sl Soares ova	1.47e-03
C 25	20	4.3	421	66	AA235526	zt30e09.sl Soares ova	1.47e-03
C 26	20	4.3	423	3	AA112480	zn68a03.sl Stratagene	1.47e-03
C 27	20	4.3	424	15	AA191137	zpb6c11.sl Stratagene	1.47e-03
C 28	20	4.3	466	79	BMAA41597	MBL2S3J2A4T3 JHU96SL-B	1.47e-03
C 29	20	4.3	466	67	AA241597	MBL2S3J2A4T3 JHU96SL-B	1.47e-03
C 30	20	4.3	503	22	AA211552	zn55f02.sl Stratagene	1.47e-03
C 31	19	4.1	360	25	N58571	yy55c12.sl Soares fet	2.53e-02
C 32	19	4.1	362	82	HS1147808	nc38f10.sl NCI CGAP P	2.53e-02
C 33	19	4.1	386	30	AA063772	ml42d06.sl Stratagene	2.53e-02
C 34	19	4.1	404	18	AA201435	LD04428.5prime LD Dro	2.53e-02
C 35	19	4.1	413	30	AA061856	ml33g02.sl Stratagene	2.53e-02
C 36	19	4.1	444	50	AA144720	mf68c01.sl Stratagene	2.53e-02
C 37	19	4.1	444	32	AA080042	mm33h09.sl Stratagene	2.53e-02
C 38	19	4.1	470	26	N76192	yz30f08.sl Soares mul	2.53e-02
C 39	19	4.1	509	101	MMAA59343	va50d07.sl Soares mou	2.53e-02
C 40	19	4.1	511	93	MM1170348	va10d02.sl Soares mou	2.53e-02
C 41	19	4.1	511	72	AA253994	va10d02.sl Soares mou	2.53e-02
C 42	19	4.1	531	31	AA065659	ml72g10.sl Stratagene	2.53e-02
C 43	19	4.1	554	52	AA184557	ml50b08.sl Stratagene	2.53e-02
C 44	19	4.1	565	18	AA202414	LD02586.5prime LD Dro	2.53e-02
C 45	19	4.1	646	15	AA191607	zq43e07.sl Stratagene	2.53e-02

## ALIGNMENTS

RESULT	1	AA203500	874 bp	EST	24-JAN-1997
LOCUS	zx58h07.r1 Soares fetal liver spleen INFLS sl Homo sapiens CDNA				
DEFINITION	clone 446749 5' similar to contains element MSRI repetitive element				
ACCESSION	AA203500				
NID	g1799268				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	1 (bases 1 to 874)				
AUTHORS	Eukaryotae; Mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
	Hillier, L., Clark, N., Dubouque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, I., Soares, M., Tan, F., Trevisakis, E., Waterston, R., Williamson, A., Wohlmann, P., Wilson, R.				
TITLE	The WashU-Merck EST Project				
JOURNAL	Unpublished (1995)				
COMMENT	Contact: Wilson RK				





```
/tissue_lib="LANL flow sorted chromosome 5 library in  
M13mp18"
```

ORDINARY

complement (2... 26)

RESULT 10  
 ID BMAA41559 standard; RNA; EST; 308 BP.  
 AC AA241559;  
 NI G1870877  
 DT 09-MAR-1997 (Rel. 51, Created)  
 DT 09-MAR-1997 (Rel. 51, Last updated, Version 1)  
 DE MBL2SJ0H3T3 JHU96SL-BmL2 Brugia malayi cDNA clone L2SJ0H3 5'.  
 KW EST.  
 OS Brugia malayi  
 OC Eukaryotae; mitochondrial eukaryotes; Metazoa; Nematoda;  
 OC Secernentea; Spirurida; Filarioidea; Onchocercidae;  
 OC Brugia.  
 RN [1]  
 RP 1-308  
 RA Blaxter M.L., Waterfall M., Daub J., Lizotte M., Baron L.,  
 RA Jones S.J.,  
 RA "genes expressed in adult female Brugia malayi";  
 RL Unpublished.  
 CC Contact: Blaxter ML Institute of Cell, Animal and Population Biology  
 CC Biology University of Edinburgh Ashworth Labs, King's Buildings,  
 CC West Mains Road, Edinburgh, EH9 3JT, UK. Tel: +44 131 650 6760 Fax:  
 CC +44 131 670 5450 Email: mark.blaxter@ed.ac.uk The ABI trace of this  
 CC sequence can be viewed at  
 CC http://www.sanger.ac.uk/bugia/L2S/MBL2SJ0H3T3.html Seq primer: T3.  
 FH Key Location/Qualifiers  
 FT source  
 FT 1..308  
 FT /organism="Brugia malayi"  
 FT /strain="TRS Labs"  
 FT /note="Vector: lambdaZapII; Site\_1: Eco R I (5' end);  
 FT Site\_2: Xho I (3' end); Brugia malayi is a lymphatic  
 FT filarial nematode parasite of humans. Full length cDNA was  
 FT prepared by long-range RT-PCR from mRNA from L2 larvae of  
 FT the human filarial nematode parasite Brugia malayi using  
 FT the human filarial nematode parasite Brugia malayi using  
 FT nematode spliced leader (SL, 5' end) and oligo-d(T) (3' end)  
 FT primers. The library had an unamplified titre of -1 x 10E6  
 FT per ml and -95% of clones have inserts (mean length  
 FT -900 bp). The library is available from The Filarial  
 FT Genome Project Resource Center: contact Dr. S.A. Williams,  
 FT Clark Science Center, Smith College, Northampton, MA 01063  
 FT USA phone +1 413 585 3826 fax +1 413 585 3786 email  
 FT genome@smith.smith.edu."  
 FT /clone="L2SJ0H3"  
 FT /clone.lib="JHU96SL-BmL2"  
 FT /sex="mixed"  
 FT /dev\_stage="mosquito derived, second stage larvae (L2)"  
 FT /lab\_host="E. coli XL1-Blue MRF-"  
 FT <1..>308  
 FT QY Sequence 308 BP; 93 A; 65 C; 70 G; 80 T; 0 other;  
 FT mRNA  
 FT 134 gtacgtatcaatggtcagcaagaacacgag 165  
 FT ||||| ||||| ||||| ||||| |||||  
 FT QY 401 GTACGTGTCATGTGCGAGCAAAATTAATTAGG 432  
 FT ||||| ||||| ||||| ||||| |||||  
 FT Query Match 4.3%; Score 20; DB 79; Length 308;  
 FT Best Local Similarity 81.3%; Pred. No. 1.47e-03;  
 FT Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 DB 134 gtacgtatcaatggtcagcaagaacacgag 165  
 FT ||||| ||||| ||||| ||||| |||||  
 QY 401 GTACGTGTCATGTGCGAGCAAAATTAATTAGG 432  
 FT ||||| ||||| ||||| ||||| |||||  
 RESULT 11  
 LOCUS BMAA41559 308 bp mRNA EST 05-MAR-1997  
 DEFINITION MBL2SJ0H3T3 JHU96SL-BmL2 Brugia malayi cDNA clone L2SJ0H3 5'.  
 ACCESSION AA241559  
 NID G1870877  
 KEYWORDS EST.  
 SOURCE Brugia malayi.  
 ORGANISM Brugia malayi  
 Eukaryotae; mitochondrial eukaryotes; Metazoa; Nematoda;  
 Secernentea; Spirurida; Filarioidea; Onchocercidae;  
 Brugia.  
 REFERENCE 1 (bases 1 to 308)  
 AUTHORS Blaxter, M.L., Waterfall, M., Daub, J., Lizotte, M., Baron, L. and  
 Jones, S.J.  
 Genes expressed in adult female Brugia malayi  
 Unpublished (1996)  
 Contact: Blaxter ML  
 Institute of Cell, Animal and Population Biology  
 University of Edinburgh  
 Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9  
 3JT, UK.  
 Tel: +44 131 650 6760  
 Fax: +44 131 670 5450  
 Email: mark.blaxter@ed.ac.uk  
 The ABI trace of this sequence can be viewed at  
 http://www.sanger.ac.uk/bugia/L2S/MBL2SJ0H3T3.html  
 Seq primer: T3.  
 Location/Qualifiers  
 1..308  
 /organism="Brugia malayi"  
 /strain="TRS Labs"  
 /note="Vector: lambdaZapII; Site\_1: Eco R I (5' end);  
 Site\_2: Xho I (3' end); Brugia malayi is a lymphatic  
 filarial nematode parasite of humans. Full length cDNA was  
 prepared by long-range RT-PCR from mRNA from L2 larvae of  
 the human filarial nematode parasite Brugia malayi using  
 the human filarial nematode parasite Brugia malayi using  
 nematode spliced leader (SL, 5' end) and oligo-d(T) (3' end)  
 primers. The library had an unamplified titre of -1 x 10E6  
 per ml and -95% of clones have inserts (mean length  
 -900 bp). The library is available from The Filarial  
 Genome Project Resource Center: contact Dr. S.A. Williams,  
 Clark Science Center, Smith College, Northampton, MA 01063  
 USA phone +1 413 585 3826 fax +1 413 585 3786 email  
 genome@smith.smith.edu."  
 /clone="L2SJ0H3"  
 /clone.lib="JHU96SL-BmL2"  
 /sex="mixed"  
 /dev\_stage="mosquito derived, second stage larvae (L2)"  
 /lab\_host="E. coli XL1-Blue MRF-"  
 <1..>308  
 Query Match 4.3%; Score 20; DB 67; Length 308;  
 Best Local Similarity 81.3%; Pred. No. 1.47e-03;  
 Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 Db 134 gtacgtatcaatggtcagcaagaacacgag 165  
 ||||| ||||| ||||| ||||| |||||  
 QY 401 GTACGTGTCATGTGCGAGCAAAATTAATTAGG 432  
 ||||| ||||| ||||| ||||| |||||  
 RESULT 12  
 ID ATA67387 standard; RNA; EST; 377 BP.  
 AC AA067387;  
 NI G1565720  
 DT 28-SEP-1996 (Rel. 49, Created)  
 DT 12-MAR-1997 (Rel. 51, Last updated, Version 5)  
 DE 26232 Lambda-PRL2 Arabidopsis thaliana cDNA clone 88F5T7.  
 KW EST.  
 OS Arabidopsis thaliana (thale cress)  
 OC Eukaryotae; mitochondrial eukaryotes; Viridiplantae;  
 OC Charophyta/Embryophyta group; Embryophyta; Magnoliophyta;  
 OC Magnoliopsida; Caprales; Brassicaceae; Arabidopsis.  
 RN [1]  
 RP 1-377  
 RA MEDLINE; 95148729.  
 RA Newman T., deBruijn F.J., Green P., Keegstra K., Kende H.,  
 RA McIntosh L., Ohlrogge J., Raikhel N., Somerville S., Thomashow M.,  
 RA Retzel E., Somerville C.;  
 RA "Genes galore: a summary of methods for accessing results from  
 RA large-scale partial sequencing of anonymous Arabidopsis cDNA  
 RA clones";  
 RL Plant Physiol. 106:1241-1253(1994).  
 CC Contact: Thomas Newman MSU-DOE Plant Research Laboratory Michigan





/sex="mixed"  
 /dev\_stage="mosquito derived, second stage larvae (L2)"  
 /lab\_host="E. coli XL1-Blue MRF-"  
 <1..>383

BASE COUNT 116 a 67 c 78 g 120 t 2 others  
 ORIGIN

Query Match 4.3%; Score 20; DB 67; Length 383;  
 Best Local Similarity 81.3%; Pred. No. 1.47e-03;  
 Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 136 gtacgtatcaatggtcagcaaaagcaatcagg 167  
 ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 401 GTACGTGTCATGTGCAGCAAAATTAATTAGG 432

## RESULT 15

ID BMAA41525 standard; RNA; EST; 387 BP.

AC AA241525;

NI 91870843

09-MAR-1997 (Rel. 51, Created)

09-MAR-1997 (Rel. 51, Last updated, Version 1)

MBL2SJOE3T3 JHU96SL-BmL2 Brugia malayi cDNA clone L2SJOE3 5'.

KW EST.

OS Brugia malayi

OC Eukaryotes; mitochondrial eukaryotes; Metazoa; Nematoda;

OC Secernentea; Spirurida; Filarioidea; Onchocercidae;

OC Brugia.

RN [1]

RP 1-387

RA Blaxter M.L., Waterfall M., Daub J., Lizotte M., Baron L.,

RA Jones S.J.;

RT "Genes expressed in adult female Brugia malayi";

RL Unpublished.

CC Contact: Blaxter ML Institute of Cell, Animal and Population

Biology University of Edinburgh Ashworth Labs, King's Buildings,

West Mains Road, Edinburgh, EH9 3JF, UK. Tel: +44 131 650 6760 Fax:

+44 131 670 5450 Email: mark.blaxter@ed.ac.uk The ABI trace of this

sequence can be viewed at

http://www.sanger.ac.uk/brugia/L2S/MBL2SJOE3T3.html This is the

full sequence of the cDNA clone. The polyA tail has been clipped

and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

CC and is excluded from this sequence Seq primer: r3.

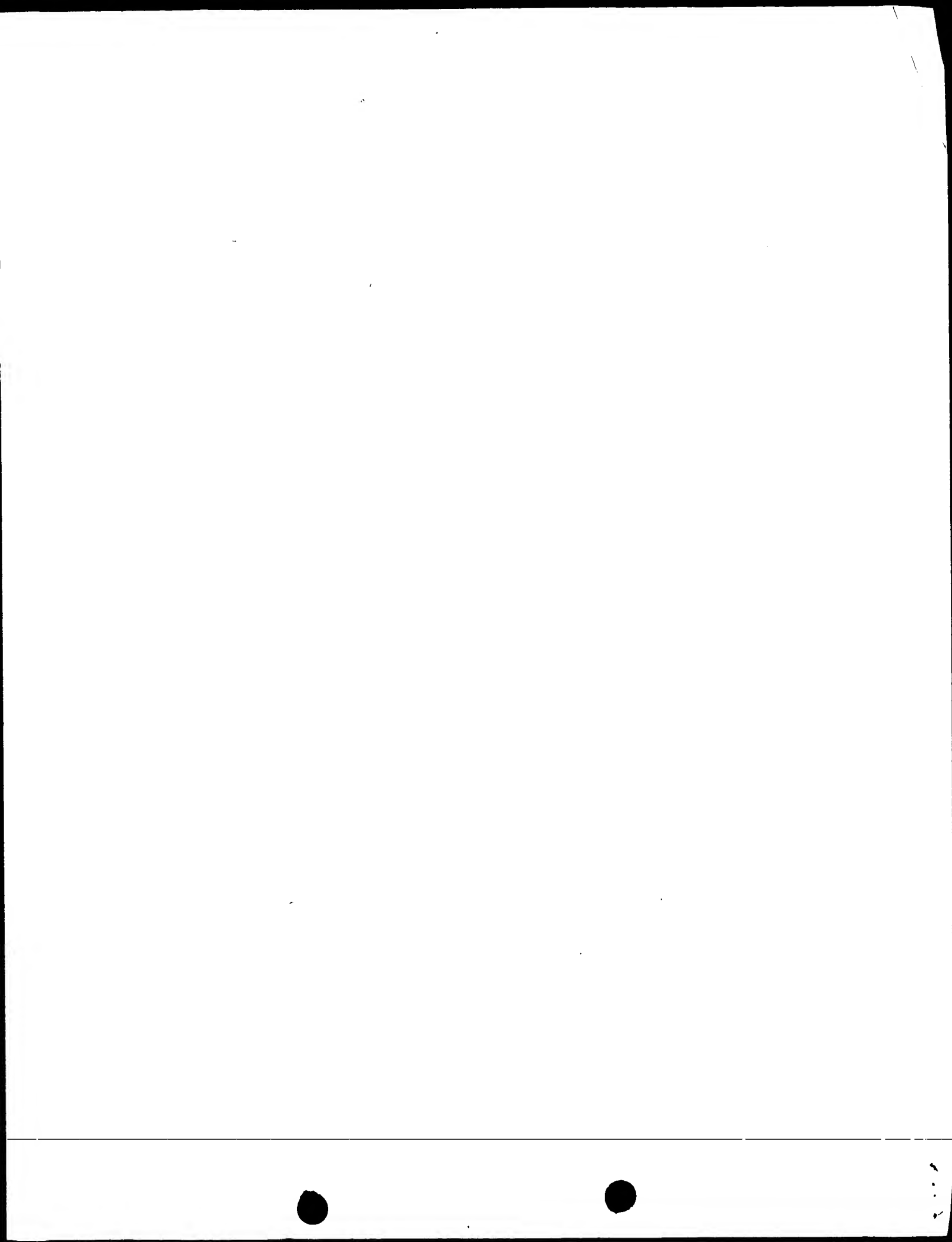
CC and is excluded from this sequence Seq primer: r3.

Db 137 gtacgtatcaatggtcagcaaaagcaatcagg 168  
 ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 401 GTACGTGTCATGTGCAGCAAAATTAATTAGG 432

Search completed: Tue Dec 2 17:14:28 1997  
 Job time : 160 secs.

source 1..387  
 /organism="Brugia malayi"  
 /strain="TRS Labs"  
 /note="Vector: lambdaZapII; Site 1: Eco RI (5' end);  
 Site 2: Xho I (3' end); Brugia malayi is a lymphatic  
 filarial nematode parasite of humans. Full length cDNA was  
 prepared by long-range RT-PCR from mRNA from L2 larvae of  
 the human filarial nematode parasite Brugia malayi using  
 nematode spliced leader (SL, 5' end) and oligo-d(T) (3' end)  
 primers. The library had an unamplified titre of ~1 x 10E6  
 per ml and ~95% of clones have inserts (mean length ~900  
 bp). The library is available from The Filarial Genome  
 Project Resource Center: contact Dr. S.A. Williams, Clark  
 Science Center, Smith College, Northampton, MA 01063 USA  
 phone +1 413 585 3826 fax +1 413 585 3786 email  
 genome@smith.smith.edu."  
 /clone="L2SJOE3"  
 /clone\_lib="JHU96SL-BmL2"  
 /sex="mixed"  
 /dev\_stage="mosquito derived, second stage larvae (L2)"  
 /lab\_host="E. coli XL1-Blue MRF-"  
 <1..>387  
 mRNA  
 SQ Sequence 387 BP; 121 A; 65 C; 78 G; 123 T; 0 other;

Query Match 4.3%; Score 20; DB 79; Length 387;  
 Best Local Similarity 81.3%; Pred. No. 1.47e-03;  
 Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;



\*\*\*\*\*

\*\*\*\*\*  
M P S R E H  
\*\*\*\*\*  
(TW)

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

MPsrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 2 17:14:49 1997; MasPar time 403.58 Seconds  
1191.939 Million cell updates/sec  
\*\*\*\*\*  
ular output not generated.

Title: >US-08-915-004-14  
Description: (1-438) from US08915004.seq  
Perfect Score: 438  
N.A. Sequence: 1 ATGAACAAGTGTCTGTGCTG.....ACGCACAGATATGATCTGCA 438  
Comp: TACTTGTTCACGACGAC.....TCGGTGTCTATACATAGACT

Scoring table:  
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 362067 seqs, 549138275 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database:  
1: BCT 2: FUN 3: GEN1 4: GEN2 5: HTG1 6: HTG2 7: HUM 8: INV  
9: OCT 10: MAM 11: VRT 12: PLN 13: PRO 14: ROD 15: SYN 16: UNC  
17: VTR  
genbank99

Database:  
18: BCT1 19: BCT2 20: BCT3 21: BCT4 22: BCT5 23: BCT6 24: BCT7  
25: BCT8 26: BCT9 27: BCT10 28: BCT11 29: GEN1 30: GEN2  
31: GEN3 32: HTG1 33: HTG2 34: HTG3 35: INV1 36: INV2 37: INV3  
38: INV4 39: INV5 40: INV6 41: INV7 42: INV8 43: INV9 44: INV10  
45: INV11 46: MAM1 47: MAM2 48: MAM3 49: VRT1 50: VRT2 51: VRT3  
52: VRT4 53: PAT1 54: PAT2 55: PAT3 56: PAT4 57: PAT5 58: PHG  
59: PLN1 60: PLN2 61: PLN3 62: PLN4 63: PLN5 64: PLN6 65: PLN7  
66: PLN8 67: PLN9 68: PLN10 69: PLN11 70: PRI1 71: PRI2  
72: PRI3 73: PRI4 74: PRI5 75: PRI6 76: PRI7 77: PRI8 78: PRI9  
79: PRI10 80: PRI11 81: PRI12 82: PRI13 83: PRI14 84: PRI15  
85: ROD1 86: ROD2 87: ROD3 88: ROD4 89: ROD5 90: ROD6 91: ROD7  
92: ROD8 93: STR 94: SYN 95: UNA 96: VRL1 97: VRL2 98: VRL3  
99: VRL4 100: VRL5 101: VRL6 102: VRL7 103: VRL8 104: VRL9  
105: VRL10  
genbank-new3

Database:  
106: BCT 107: GEN1 108: GEN2 109: HTG1 110: HTG2 111: INV  
112: MAM 113: VRT 114: PHG 115: PLN 116: PRI1 117: PRI2  
118: ROD 119: SYN 120: UNA 121: VRL

Database:  
u-emb150.99  
122: part1

Statistics: Mean 10.182; Variance 3.871; scale 2.630

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
C 1	28	6.4	215 57	128278	Sequence 5 from patent	7.95e+05
C 2	26	5.9	215 57	128278	Sequence 5 from patent	2.36e+03
C 3	24	5.5	7095 92	RRMAP1B5	R.norvegicus mRNA for	6.05e-02
C 4	22	5.0	1245 61	CREPHOSRI	C.reinhardtii phospho	1.31e+00
C 5	22	5.0	1256 47	OCSPI7GN	O.cuniculus spl7 gene	1.31e+00
C 6	22	5.0	1388 118	MMTNFR2A	M.musculus tumor necr	1.31e+00
C 7	22	5.0	1388 86	MMTNFR2A	M.musculus tumor necr	1.31e+00
C 8	22	5.0	1505 89	MUSMTNFR2	Mouse tumor necrosis	1.31e+00
C 9	22	5.0	3796 90	MUSMTNFR1	Murine tumor necrosis	1.31e+00
C 10	22	5.0	8818 86	MMNAD1B	Mouse MAP1B mRNA for	1.31e+00
C 11	22	5.0	9416 83	HUMSEOX	Human microtubule-ass	1.31e+00
C 12	22	5.0	17013 20	EAAMASL	E.amylovora (Ea7/74)	1.31e+00
C 13	21	4.8	683 61	CRU19490	Chlamydomonas reinhar	5.68e+00
C 14	21	4.8	1671 63	MZEGSTIA2	Maize glutathione S-t	5.68e+00
C 15	21	4.8	1913 48	SSACHRM	Porcine DNA for musca	5.68e+00
C 16	21	4.8	2356 45	SPASUBA	S.purpuratus nuclear	5.68e+00
C 17	21	4.8	2538 46	BTU08286	Bos taurus muscarinic	5.68e+00
C 18	21	4.8	37617 24	MLB1790G	M.leprae genes rpll,	5.68e+00
C 19	21	4.8	144861 98	HU43400	Human herpesvirus-7 (	5.68e+00
C 20	21	4.8	201239 109	HS435D1	Human DNA sequence **	5.68e+00
C 21	21	4.8	201239 5	HS435D1	Human DNA sequence **	5.68e+00
C 22	20	4.6	69 56	II5665	Sequence 2 from patent	2.33e+01
C 23	20	4.6	105 56	II4734	Sequence 13 from pate	2.33e+01
C 24	20	4.6	990 74	HSB2BH	H.sapiens H2B/h gene.	2.33e+01
C 25	20	4.6	1165 80	HUMGT212A	Homo sapiens GT212 mR	2.33e+01
C 26	20	4.6	1192 41	DROANNX	D.melanogaster annexi	2.33e+01
C 27	20	4.6	1330 122	BPPT	B.pendula mRNA encodi	2.33e+01
C 28	20	4.6	1330 115	BPPT	B.pendula mRNA encodi	2.33e+01
C 29	20	4.6	1871 67	SOYICL1A	Glycine max glyoxysom	2.33e+01
C 30	20	4.6	2205 68	WTRUBIAA	T.aestivum ribulose-1	2.33e+01
C 31	20	4.6	2490 68	WTRUBIAA	Trichoderma harzianum	2.33e+01
C 32	20	4.6	2619 111	DMANX	D.melanogaster anxx g	2.33e+01
C 33	20	4.6	2619 39	DMANX	D.melanogaster anxx g	2.33e+01
C 34	20	4.6	3022 40	DRO14DC22	Drosophila melanogast	2.33e+01
C 35	20	4.6	3865 18	ACNVFDGK	Azotobacter thiooxid	2.33e+01
C 36	20	4.6	3899 19	BNRRTEAB	Bacteroides thetaiota	2.33e+01
C 37	20	4.6	5208 89	MUSPC6B	Mouse mRNA for PC6B,	2.33e+01
C 38	20	4.6	5588 41	DSPTSH1	D.subobscura DNA homo	2.33e+01
C 39	20	4.6	13356 104	RCU50550	Rat cytomegalovirus m	2.33e+01
C 40	20	4.6	30911 38	CEU06C7	Caenorhabditis elegans	2.33e+01
C 41	20	4.6	31001 80	HUMCOL2A12	Human pro-alpha type	2.33e+01
C 42	20	4.6	41611 74	HSF77D12	Human DNA sequence fr	2.33e+01
C 43	20	4.6	66990 40	DRO14P	Drosophila melanogast	2.33e+01
C 44	20	4.6	88971 118	MMU58105	Mus musculus Btk locu	2.33e+01
C 45	20	4.6	185790 110	CEY37D8	Caenorhabditis elegans	2.33e+01

ALIGNMENTS

RESULT	1	128278	215 bp	DNA	PAT	30-OCT-1996
LOCUS		Sequence 5 from patent	US 5569830.			
DEFINITION		Sequence 5 from patent	US 5569830.			
ACCESSION		128278				
NID		91819054				
KEYWORDS		Unknown.				
SOURCE		Unknown.				
ORGANISM		Unclassified.				
REFERENCE		1 (bases 1 to 215)				
AUTHORS		Bennett, A., Labavitch, J.M., Powell, A. and Stotz, H.				
TITLE		Plant inhibitors of fungal polygalacturonases and their use to control fungal disease				
JOURNAL		Patent: US 5569830-A 5 29-OCT-1996;				

FEATURES  
source  
Location/Qualifiers  
1..215  
/organism="unknown"  
BASE COUNT 15 a 8 c 25 g 26 t 141 others  
ORIGIN

Wed De

		Query Match Best_Locals_Matches	Dc	39	Cp	319	Dd	99	Cp	259	Dd	158	Cp	199		RESULT	LUCUS	DEFINITION	ACCESSION	NID	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES	Source	BASE COUNT	ORIGIN	Query Match Best_Locals_Matches	Dc	1 m	Qy	219 C	61 Y	278 T	Dd	121 g	Qy	337 C	RESULT	LOCUS	DEFINITION	ACCESSION	NID	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL MEDLINE REFERENCE
--	--	------------------------------------	----	----	----	-----	----	----	----	-----	----	-----	----	-----	--	--------	-------	------------	-----------	-----	----------	--------	----------	-----------	---------	-------	---------	----------	--------	------------	--------	------------------------------------	----	-----	----	-------	------	-------	----	-------	----	-------	--------	-------	------------	-----------	-----	----------	--------	----------	-----------	---------	-------	---------------------------





TITLE Cloning and expression of cDNAs for two distinct murine necrosis factor receptors demonstrate one receptor is species specific  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 88, 2830-2834 (1991)  
MEDLINE 9118785  
FEATURES Location/Qualifiers  
source 1..1505  
/organism="Mus musculus"  
sig\_peptide 41..106  
CDS 41..1455  
/codon\_start=1  
/product="murine tumor necrosis factor receptor 2"  
/db\_xref="PID:g19828"  
/translation="MAPAALWALVLFELQLWATGHTVPAQWLTTPYKPEGYECQISQ  
EYDRKAMCCAKPPGOYVKHFNKTSDTVCADCEASMYTVVNFRTCLSCSSCT  
TDVEIRACTKQNRVACAEAGRYCALKTHSGSCRCQMRSLKCGPGEVASSRAPNG  
VLCKACAPGTFSDTSSDYCRPHRISILAIPEGNASTDAVCAPESFLIPRTLIY  
SQPEPTSRQPLDQPGSPQPSILTSIGSTPIIEQSTKGGISLPIGLIVGTVSLGLM  
LGLVNCILVORKKKPSCLORDAKVPHVDEKSDQAVGLEQOHLITTHVNTICVNVCS  
SASAGDRRAPPGGHPQARMAEAOQFARASRISDSHSGHGHHTVNTICVNVCS  
SDHSSQCSOASATVGDPAKPSAPKDEQVPFSQECPSQSPCTETTLQSHKPLP  
LGVPDMGMKPSQAGWFDQIAKVA"  
347 a 459 c 408 g 291 t  
ORIGIN

Query Match 5.0%; Score 22; DB 89; Length 1505;  
Best Local Similarity 64.1%; Pred. No. 1.31e+00;  
Matches 50; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

Db 193 tcagatgtgctgtaagtctctcctgccaatatgtgaacatttctgcaacaagac 252  
QY 111 TCAGCTGTGTGTGACAAATGCTCTCTGCTACCTACCTAAACAACACTGTACAGCAA 170

Db 253 ctcggacacccgtgtgtgc 270  
QY 171 GTGGAAGACCGTGTGCGC 188

RESULT 9  
LOCUS MUSTNFR1 3796 bp mRNA ROD 01-AUG-1991  
DEFINITION Murine tumor necrosis factor I receptor (TNFR-1) mRNA, complete cds.  
ACCESSION M59378  
NID g202094  
KEYWORDS tumor necrosis factor receptor.  
SOURCE Mus musculus lymphoid cDNA to mRNA.  
ORGANISM Mus musculus  
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Euthera; Rodentia; Sciurognathi; Myomorpha; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 3796)  
Goodwin, R.G., Anderson, D.M., Jerzy, R., Davis, T., Brannan, C.I., Copeland, N.G., Jenkins, N.A. and Smith, C.A.  
Molecular cloning and expression of the type 1 and type 2 murine receptors for tumor necrosis factor  
Mol. Cell. Biol. 11, 3020-3026 (1991)  
JOURNAL 91246168  
MEDLINE  
FEATURES Location/Qualifiers  
source 1..3796  
/organism="Mus musculus"  
/cell\_line="7B9"  
/cell\_type="helper T-cell"  
/tissue\_type="lymphoid"  
/map="Chromosome 4"  
/gene="TNFR-1"  
/codon\_start=1  
/db\_xref="PID:g202095"  
/translation="MAPAALWALVLFELQLWATGHTVPAQWLTTPYKPEGYECQISQ

sig\_peptide 43..108  
CDS 43..1457  
/gene="TNFR-1"  
/codon\_start=1  
/product="tumor necrosis factor receptor 1"  
/db\_xref="PID:g202095"  
/translation="MAPAALWALVLFELQLWATGHTVPAQWLTTPYKPEGYECQISQ

EYDRKAMCCAKPPGOYVKHFNKTSDTVCADCEASMYTVVNFRTCLSCSSCT  
TDVEIRACTKQNRVACAEAGRYCALKTHSGSCRCQMRSLKCGPGEVASSRAPNG  
VLCKACAPGTFSDTSSDYCRPHRISILAIPEGNASTDAVCAPESFLIPRTLIY  
SQPEPTSRQPLDQPGSPQPSILTSIGSTPIIEQSTKGGISLPIGLIVGTVSLGLM  
LGLVNCILVORKKKPSCLORDAKVPHVDEKSDQAVGLEQOHLITTHVNTICVNVCS  
SASAGDRRAPPGGHPQARMAEAOQFARASRISDSHSGHGHHTVNTICVNVCS  
SDHSSQCSOASATVGDPAKPSAPKDEQVPFSQECPSQSPCTETTLQSHKPLP  
LGVPDMGMKPSQAGWFDQIAKVA"  
109..1454  
/gene="TNFR-1"  
/codon\_start=1  
/product="tumor necrosis factor receptor 1"  
BASE COUNT 863 a 1052 c 981 g 900 t  
ORIGIN Chromosome 4.

Query Match 5.0%; Score 22; DB 90; Length 3796;  
Best Local Similarity 64.1%; Pred. No. 1.31e+00;  
Matches 50; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

Db 195 tcagatgtgctgtaagtctctcctgccaatatgtgaacatttctgcaacaagac 254  
QY 111 TCAGCTGTGTGTGACAAATGCTCTCTGCTACCTACCTAAACAACACTGTACAGCAA 170

Db 255 ctcggacacccgtgtgtgc 272  
QY 171 GTGGAAGACCGTGTGCGC 188

RESULT 10  
LOCUS MMAP1B 8818 bp RNA ROD 12-SEP-1993  
DEFINITION Mouse MAP1B mRNA for MAP1B microtubule-associated protein.  
ACCESSION X51396  
NID 952999  
KEYWORDS MAP1B gene; microtubule-associated protein.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Euthera; Rodentia; Sciurognathi; Myomorpha; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 8818)  
Cowan, N.J.  
Direct Submission  
Submitted (09-JAN-1990) Cowan N.J., Dept. of Biochemistry, New York University Medical Center, 550 First Avenue, New York, NY 10016  
REFERENCE 2 (bases 1 to 7620)  
Noble, M., Lewis, S.A. and Cowan, N.J.  
The microtubule binding domain of microtubule-associated protein and tau  
MAP1B contains a repeated sequence motif unrelated to that of MAP2  
J. Cell Biol. 109 (6 Pt 2), 3367-3376 (1989)  
JOURNAL 90094539  
MEDLINE  
FEATURES Location/Qualifiers  
source 1..8818  
/organism="Mus musculus"  
/strain="Swiss Webster"  
/dev\_stage="5 day old"  
/tissue\_type="brain"  
/clone\_lib="lambda GEM"  
/note="primary transcript"  
/db\_xref="PID:g53000"  
/db\_xref="SWISS-PROT:P14873"  
/translation="MATVVEATEPESSIGNPAASTSPSLSHRFLDSKFLYLVVVG  
ETVTEHLRAIGNIELGRISWDNLIECNLDQELKLVFSRHSARFSEVPQKILHH  
RSDVLETVLINPSDEAVSTEVRLMTDAARKLLVLTQCFENTGELILOSGSFQ  
NFIEFTDQIEGELLSTHPANKASLTLCPEEGDKNSLDRLNQDFINIKNSAS  
ILPEMEGLSEFEYELISEVSEVPFDELPDITSGFLKLSKPCYIFPGGSDSALFA  
VNGFNLLNGGSKERKSCFWKLIRHLDVDSLTHIGDNLPGINSILQRIAELEE  
RSQGSTNSDMKNLISDPLGVFLNPNENKDPENPKMKRSITEACFTIQLYNKLS  
MKPEPLERSVGNITIEPVILFQKMGVGLKLVNPNVPSKSKEMQYFMQQTCTNKDAE

LILPNQOEVDIPISLYLTVSSLLVHPHPANPAEKIIRVLFGNSTQYVILBGLKXHL  
 DFLKQPLATQKDUIGTOPPPVQWKLQKQADRESLQAPATKPVASQSVKRESKEPT  
 EVTKISOVERTKPKSEKVKLVKKDKQVKRTSEKSPVTEKSEVKEQSPVKAEBQ  
 ATSPKQVKKDKVVKKEIKTLEKKEEKPKVEVVKKBDTKPLKDKPKSEVKEKEI  
 KKEIKPEKRELKKEVTKLTKDKAKKEKKEKVEKKEKPKKIKKIDKIKKS  
 TPOSDTKPKSALQPKVAKVEESTKEPILAAGLKGKVKVKKKEGTTEAAATVOLT  
 AANTAAVVAAGAASPVKVELEASERSLWSSPDOLTKDFELKAEETDVAKDJKPQLE  
 LIDEBEKLKETOQGEAYVLOKTEVYSKGSASPOBGIITTEGEGECQTPPEELPEVVK  
 QGVQDILFEGEGESESSESTDYEEAEKAEETAEPEEEDGDNASGASKHSPTEDE  
 ESAKAEADHGLKKEKRESVGGDRAEEMADVDLEKGAEOSEEEEGEDKRAEDAREEG  
 YGSDTAEADYMAVADKAAEAGVTEQYQGLTSAKQPGIQSPREPAISIHDETLPL  
 GGSEIATSDENRDQDEEPTATSGTYGTITLISEPSTPMDSEMTSPROVMSDETNL  
 EETSPQEPVNIITKYESLVSQYKSPAVASFNGLSEGSKTDATGDKQVNAASATIS  
 PRSMEDEKFSALDACYSEBEKLKASALDIKQVSDERLSPAKSPSISPSPSPPI  
 EKTPLGERSVNSLTFNEIKVSEBEKELKASALDIKQVSDERLSPAKSPSISPSPSPPI  
 SVTSGAGETVYQSPIDERSKSHLPIEVSENAQVPSFEPSEAKDENERASLSPMDPEP  
 VPQSESEKVKVPSLPLGLGSESGTFIPIDEDFEPKTKSDVETMSQSALADRLKGLG  
 ESVPTDITGTSQFGSEKDEKTKMGISIEGTVSDKSAIPVDEGVAEDTYSHMEGVASVSTL  
 VATSSPEPTDVSATKSHAEVYSPHSTVEDDLSVSVVOIPTTFQTEMSPSKCEPC  
 RWSJLPPDEPTAKSPYQVODHSESOSSMSIFGQSESPHSPMDFRSQSDPHTL  
 GASVLIHITENGFEVYSCDQDSSLSHKIPPTPEEYSPQDNDLSLISVSOVEASP  
 STSPASJPSQIASPLEQDLSLSDVPPRMSLYLASKEVOSLEGELKSPKSDISPLT  
 PRSESPJSGESDTSAPKATAFAAHQASSPIDAATAEYPFRSMFLDFTMOHHLA  
 LNRLDITTSVEKDSQKGPEDYNYATKQPDENAGSPDEEDYDESQETKTRTHDVYVY  
 YKERTKTSKPCDQSGSYETIETKTKTPEDGAGYCEITETKTRTPDEGGYSYKEIEX  
 TTRTPYEVSTNKTPEBRSRRLCDLSNGYDIDTQDGHTLGDGCSVYETTEKIFSPES  
 ESYSYETSNKTTPESDPAKYETMEKIKTKPQASTSYETSDRCVYTEKKSPSPARO  
 DVDLCVLVSCSEPKHPTLSPSFINPNLPHFAGEETRESEKPLTQSGGAPPPSGGK  
 QQRQCDTEPTPYSSESASFQSDSPLEPEECPSTADANDISESETITPIDKTVT  
 YKRMQDPPAPMQDPSPPRHPDVSMDPDALAVQNLGKAVKDLKESEIKTKPQTKT  
 KSSGPKVKGDGSKPLAASPKGALKESDKSVSPKPKSEYKRAIKTITTPYVKA  
 TRREEDKTKTNAANASAKSATITGPTTKTAKSTVPPGLVPLVYDLCYTPNHS  
 SKNVDEYFFYKRRVSYNMDPAAEPSPRAVLDALECKGAOWGSMNQVLLTPTHDS  
 ENKREVOETHEKQOOLNIVYLASSSTVVMODESEPAKTEI"

/translation-"MATVYENATEPPSPGSGSIANPAASTSPSLSHRFLDSKRYLLVTVVG  
EIVTEHRLRAIGLNGELIRSDMTLNIENLDQELKLVFSRHSRASTPEVPVQSGOLHFF  
RSVDVLTVINPSDEAVSTVRYMLTIDAAHKLKLVFQGFENTGELIGLQSLQSLU  
NTEITDQELGELLSTHPANKASDITPEEGDGKNSLDNRHNLQNTIKNLKNGAS  
ILPEMGGLSEFYTESVSEVSPSDIILEPTSGGFLKSKPCCYIPFGGGRDSALFA  
VQSGNTLNGGSRKSCFWKLRLHJDRDSILLTHIGDNLPGINSMLQRLIAELDEE  
VQSGNTLNGGSKNLI SPLDGVVFLVNPENLKNPEPNIKKMRSTIERACFTLYQVNLKLS  
MKPEPLRVSGVNTLDPYLLFKMGVGVKLEMYVLYNPVSKSKEMQYFMQWGTGKNKRAE  
FILPNQGVDFLSTSVSLNIVLWHPNPAEKIRVLPFGNSTQYILGSLGELKHLH  
DELKQPLATQKDLTGVPVYVQVKTLKORADRESUKPAAKPLPSKVSRSKSEKTEPP  
EVTNVNHRKPKTKVKEKVKMDKPKVITETKPSVTEKEVPEKSEEPSYKVAEBK  
ADTVKAPAKETKKEKVKEDKPEKKEKPKVEKAKEDKTPIKEEPKPKVEVK  
EVKEKIEKKEKPKVEKKEPPEKVEKVEKKEKVEKKEKKEKIKLPLDKAB  
KSTPSLEAKKPAALKVPKPEKSVKVDKSAAGKPKKEGKIIVIKKEGKAASVAA  
VGTGATTAAMAAGTAAIGPAKLELAEPSRLMSPEDLTNDKFELEAEVDVTKDKLP  
QELAELEAEVAVLQKREVTYVLTQKPAESPDGELITLTHEGECEBOTPELEP  
VEKGVDDIIEKDEDEGAPESSEGTGYEKAATEEAEEPEEGEEHVCVASKHSPT  
DEESAKADAPALIEKRESVAGSDRADAEDDEALIEKGAQSEEADEEADKAEAD  
EESTYKEEMADYVMAVVDKAAEGAGAEQYGLTPTTKOLGQASGPRSEPAASIHDE  
TLPGSESAATSDNEEDRQDEEETATSGYQSTIIEISSEPTMDMSPTDVMDSDE  
TNNEETSEQFANITYEESLYSQEYKPADVPLNGSEKTDATQKQDYNASA  
STSPSPMEEDKFRSALRDYKAEVKAATLTDIKDSISAVSEKVSPEKSLSPAS  
PPSLSEKTLTGERSVNSLTLPNEIKVSAEAVAPVSEPTQVVEVHEHCAKSPDKLTV  
VSVSQSVTSGAGHTPYQSDTGHSEHUTBEVIEKPAVPYSPFSDAKDENRASVS  
PMDEPVDSESTIEKVLSPALSPPLIGSEVESFLSADDKASGAESEPEKSGQ  
GSPDQVSPVSEWTSNLYQDKQEGKSTDEAPIKEDQGEKTKDDVEMSSOPALADE  
RKLGDVSPYQIDVYSGQSEKEDKMSISEGTVSKDSATPYDGEVAETIYSHMEGVASV  
STASVASISFPDPTIDDVSLHJADVHPSHTEVDDLSVSVVGTPTIDTEMSPSK  
EPCPRMNSIPDPDPFKTAKSTVQMDHRSGQSMSEIEFQGESPEQSLAMDVSRQSPD  
HPVAGVLHVLTENGPTEDVYSDPMDSKSLHKKIPMEFSPSTQDNDSLELISVSQ  
EASPSTSAHTPSQSLPQEDTSDVAPPRDMSLYASLTKSEKVSQLEKSLSPKSDI  
SPITPRESPLSTPDSOTSAVKTEKTHCSSSPPIDAASEPTVGFRAVLFLDTMQ  
HLIALLNRDITGLKDGSGKTPGFSYAQKPEETRSPEDEYDYESYKETRSD  
VGGYVYKELERTKSPSDGSGYETIGKTKTPEDGDYSYETIKETRTPEEGDYSYD  
ISEKTTYSPVEVSGYSEKTERSRLLCDTISNGYDSDGHTGLDPSYSYETKTKITS  
FPSEYSEYETSKTRTDPDSTYCTAEKTRTPQOASYSYETSDLYTAEBKKSPP  
EABQDYDLGLVSCYEKHPKHELSPFINPNLFWAFASBEPTSEKPLTQSGCAPPP  
PGKQKRODDEPPTPEPTEPTECPSTSEKPLTGAALDKDESESTETPTD  
KTVYVGHMPPPPVODRSPSPRPDPVSNWDPDEALAEONLGAALDKDESEKTKTKK  
GTTKSSSPVSKDSKSPPLAASPKAGLKSEKSRVASPKKSEKVEAKAPPTTP  
EYKAAEGSEKDETNAANASAKSAKTATAGPQTKTKTSVAVPGCLPYLDLCYIP  
THNSKNRNVGFFKRVSSYVVGSDNPAAEPEPSRAVLDALEKGAQSGNSMGVTLIP  
THDSVYRWYEWOTHEKODUINLVASLSTVWODESPACKITL

[illegible]



MEDLINE 95319333  
REFERENCE 2 (bases 1 to 9538)  
AUTHORS Geider K.K.  
TITLE Direct Submission  
JOURNAL Submitted (01-MAR-1994) K. Geider, MPI fuer medizinische Forschung,  
Jahnstr 29, 69028 Heidelberg, FRG  
REFERENCE 3 (bases 1 to 17013)  
AUTHORS Geider K.K.  
TITLE Direct Submission  
JOURNAL Submitted (06-DEC-1994) K. K. Geider, MPI fuer medizinische  
Forschung, Jahnstr 29, 69028 Heidelberg, FRG  
FEATURES  
source  
1..17013  
Location/Qualifiers  
/organism="Erwinia amylovora"  
-35\_signal  
507..512  
-10\_signal  
530..535  
RBS  
1048..1051  
/gene="amsG"  
CDS  
1058..2491  
/gene="amsG"  
/codon\_start=1  
/product="UDP-galactose-lipid carrier transferase"  
/db\_xref="PID:g600427"  
/transl\_table=11  
/translation="MREIEFTFKGLIRLSALSLDIPFNIALALAIVLINGFPGEIL  
TGIPQHELDLKIATIHLLSVICGVFWVRLRHRYRKPFELKEVFTILIFSVIDL  
SVSALSKEWLSWILWLLSMAMPFGACVRLNRKLLKWKQSIIGSGKNAOE  
AQWALQSREMGEFDIAFYDVGDSOTALEFGVPLKEEQQLSLVDSQTFIVAVEY  
EQSQDRWLKLNATHNRSVSPSLRGVPLGTDMAYIISHEWMLRNLNSLAKHS  
SFLKATDLVGLASIIITLLPALVILIFVRSDDGAPYIGHERVGRDGRKFKLKLR  
SMVANSKEVELEVETDPVARAENDEFKLNKDPRITRIGHFIRKTSLELPEQLMNVY  
RQMSLVGRPVIEALRYAGDVDFYFMAKPGMTGLMQVSGRNDVSYETRYFDSWY  
VKNWSLWMDIALFKTIGVILKRDGAY"  
2600..2604  
/gene="amsH"  
RBS  
2613..3746  
/gene="amsH"  
/codon\_start=1  
/product="outer membrane protein"  
CDS  
/db\_xref="PID:g600428"  
/transl\_table=11  
/translation="MIITKLIPLMVPAAALLSGCTIVPGNHLSTMCKDQVVEQDSDF  
LDIKYVIFPLPSLVERMKPPVQAQANATLQRELQNYEYRIGVDVLMVYVNDHPE  
LTTAGQVRSKQYTYTGEVTTSGQAITNPLTITLDAINAAAGLTATADNRNVLTTHD  
VQNVASFSKQYTYTGEVTTSGQAITNPLTITLDAINAAAGLTATADNRNVLTTHD  
GROVPSLQALMONGDPSQNHLLYPGDILYVRNDDLVFVYMGVEVKQATLKMDRSGM  
TISEALGSAQGDQVADATGVFVIRPVKGNRNIANIYQNTKDAAMVMTGEFRL  
EPYDIVYTSPLTRNVRVISQLVTSIGVYDARNVQTIHKW"  
3749..3753  
/gene="amsI"  
RBS  
3762..4196  
/gene="amsI"  
/codon\_start=1  
/product="acid phosphatase"  
CDS  
/db\_xref="PID:g600429"  
/transl\_table=11  
/translation="MINSILVVCIGNICRSPTEBRLKAALPERKIASAGLKAMVGS  
ADETASIVANEHGVSLQDHVAQLTADMCROSDILVNEKKHIDLVCRIKNSVRGKTM  
LFGHWINQEIADPYKKSRADEAVYGVLENAQKWNALSR"  
4199..4202  
/gene="amsA"  
RBS  
4209..6389  
/gene="amsA"  
/codon\_start=1  
/product="membrane-associated ATP-hydrolase"  
CDS  
/db\_xref="PID:g600430"  
/transl\_table=11  
/translation="MKSRIEFPANGGEAIFELAHLLGQLLDRHWMIVAVSVLFTLM  
GTYLSIFATPIYADAMVQVEQKANTVINDISQMPNAPASAAEIEITSRMVLGK  
TVADLGLDVLQGHFLIGAGLSRIIGOKAQOIAVSRKLVPTLWDKRELSVEVDGPD  
SVTVSKDGNELFKGVQFEOHQDVTMLVNSIEADAGTRFTVSKLNNLQAIKMISNL  
VYADMGKTGVLFGYSSGDPQVSRVLDQVINYLNQNIARKSEEEKSQFIAQQL  
PDVRKLAQDAEDKLVNFRKKHDSVDMSLEAKSDSSVSITQNLALTFRREAESQLF

RBS  
CDS  
/gene="amsB"  
6478..7383  
/gene="amsB"  
/codon\_start=1  
/product="glycosyl transferase"  
/db\_xref="PID:g600431"  
/transl\_table=11  
/translation="MKDISESVVIPAYNASESIITLCLNEQSYKNEFDVIVDDKSA  
DAOKLAEVSSERYSGKLNLVSETKLNGAGNRGIDLATGDVYCFDADDEWHKD  
KQQLNLSLIERLEGQDRRIIYSQVNIIDGSEFLKMPKPVGESEHAYLFGCYG  
FIQSTIVLKREDAEIRREDERYIRHODYDLICIRADKLGKFMIAQPLANYHMVTRF  
GSOHGESVKYSLFLWDAMKPHLTRRDVYTKAYKLPLRYKMDGKSLQASLSFARYFF  
LTKNDNRNDFLKLNNKLRLTGK"  
7404..7406  
/gene="amsC"  
RBS  
7414..8541  
/gene="amsC"  
/codon\_start=1  
/db\_xref="PID:g600432"  
/transl\_table=11  
/translation="MAIYVIVSYSLVFCFFELAMINQASEAKTKILINFFLIGVFA  
LILFAGIRPGSDMDSQVFFHDFSRQGTQYQAVANIYRNFVLAWVASCF  
THEYVLTIFISFIAVSTNAMYKYPILICLSLYSAHLEINKDMNIGRLGASAF  
ATAFICSLVARNYLLALLVLTSTQSHSYTIVMIIPFFIRKYLPLVIVIASIP  
LGTIGGKPLDLSLGIYPVYGERAASYSCTNFTSPVGLAKLNIAFGAFTLYFF  
RGIMKEDRFVILLIAYSIGAARVITFSDPSIFGRGVNGLFLHTEPLLPALMLAIR  
NLLNFMLFSTIITTYLAYNTILSAPSIMGYSVAPLFRIFS"  
8695..8698  
/gene="amsD"  
RBS  
8707..9762  
/gene="amsD"  
/codon\_start=1  
/product="glycosyl transferase"  
CDS  
/db\_xref="PID:g600433"  
/transl\_table=11  
/translation="MYKLLIDIGISGGTDRVASTLSLLSNHNDVLYLSLNSGE  
PYVPDLSIROPKSSMLKFLFEIRYAKNTRPDGVMIISMGKLSVQALLSLKFLR  
KVSRLCCDHVSIETLSAAVRLKVCYGLAEKVVLTOHDKNLITSAFLKNYVVG  
NISPFHNSLNRFDFARKQNRVLAVGRLYIQKNGFLDIWKNHKGWKLIVG  
DGEKAELLEKIKKHYLESSEIIVSPSKKISSEYRSSGVIAVTSRYEGLPWLIEAKN  
YALPAIADFCKTGPAEIIRDDGYVVDYQSNELFTAQNLQIASQELRKNFAQAQWNS  
ADYGPGLILKWNIDILN"  
9767..9770  
/gene="amsE"  
RBS  
9775..10575  
/gene="amsE"  
/codon\_start=1  
/product="glycosyl transferase"  
CDS  
/db\_xref="PID:g600434"  
/transl\_table=11  
/translation="MFSVLISLYNKEKPENLQCLLESHQQLNADEIVLYVDGPVSE  
SLKAVATRWANLLPLVIVPELNLGLKALNAGLERCTHNVVARMDTDDICUPREFEK  
QISYMESHPEVVLGSAVIEFDEHKGKRLKRLPLSNNDIHEFAKRNHNCVYFRK  
DKVISAGYQHLLYMEDYLNRLMSLHPVANLPDLVLMKVRAGSDMVKNRGWNVIK  
SEVQLYRLKALKQTGFIKRTGLYFLIRTWTMLPVKVMQFLYEKDRKG"  
10652..10655  
/gene="amsF"  
RBS  
10662..12893  
/gene="amsF"  
/codon\_start=1  
/db\_xref="PID:g600435"  
/transl\_table=11  
/translation="MKRRLIITAFSTIVATAALSVSARSSEDLHGLTLKKVPPDA  
IPANDVIFSPQDVFPTMPQDFWRDKGLYIGKAGTDPTLPONLIDVFNKNGSTAL  
LSQIDLNSETLTKTFVAARGALWSASEYSMALHNDNDEQIFVYVPVKNGVSEFGRRL



Best Local Similarity 78.4%; Pred. No. 5.68e+00;  
Matches 29; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Db 1368 gaagtcgctgctgaagaccatctgcttgaac 1404  
||||| ||||| ||||| || ||||| |||||  
Qy 325 GAAGGCGCTACTTGAGATAGATTCTGCTTGAAC 361

RESULT 15  
LOCUS SSACHRM 1913 bp DNA MAM 12-SEP-1993  
DEFINITION Porcine DNA for muscarinic acetylcholine receptor III.  
ACCESSION X12712  
NID g1861  
KEYWORDS acetylcholine receptor.  
SOURCE pig.  
ORGANISM Sus scrofa  
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;  
Vertebrata; Eutheria; Artiodactyla; Suiformes; Suina; Suidae; Sus.  
REFERENCE 1 (bases 1 to 1913)  
AUTHORS Akiba, I., Kubo, T., Maeda, A., Bujo, H., Nakai, J., Mishina, M. and  
Numa, S.  
TITLE Primary structure of porcine muscarinic acetylcholine receptor III  
and antagonist binding studies  
JOURNAL FEBS Lett. 235 (1-2), 257-261 (1988)  
MEDLINE 88296835  
COMMENT library=lambda EMBL3; clone=pSpmACR7.  
Data kindly reviewed (03-OCT-1989) by Numa S.  
FEATURES  
Location/Qualifiers  
1..1913  
/organism="Sus scrofa"  
93..1865  
/note="muscarinic acetylcholine receptor III (AA 1 - 590)";  
/codon\_start=1  
/db\_xref="PID:g1862"  
/db\_xref="SWISS-PROT:P11483"  
/translation="MTLHNNNTSPLFPNISSSWIHGSPDAGLPPTVTHFGSYNISQ  
AAGNFSPPNGTSDPLGGHTIQQVFIATLGTILALVTIIGNIIIVIVAFVKNKOLKTV  
NNYFLLSACADLLIIGVISMNLFTYIIMNRWALGNLACDLWLSIDYVSNASVMNLL  
VISFDRYFSITRPLTYRAKRTTKRAGVMIGLAWVISFILWAPILFQWYFVGRKTVPP  
GECFQIFLSEPTITFGTAIAAFYMPVTIMTILYWRIVYKETEKTRELQAGLQASGTEAE  
AENFVHP7GSSRSGLSSLELQOQSLKRSARKYGRCHFWEFTKSWKPSAEQMDQDHSS  
DSWNNDAASLENSADEEDIGSETRAIYISIVIKLPGHSTILNLTPLPSSONLQVP  
EEELGTVDLERKASKLQAKMSDDGSGFQSKLPQLQESAVDTAKASDVNSVGTK  
TATPLSPFEAKTALRQITRKRKMSLIKERKAAQTLSAILLAFIITWTPYN  
IMVLNTPCDSCIPKTYNLCYWLICYINSTVNPVCYALCNKTPRTTFKMLLQCQCDKR  
KRRKQYQQRQSVIFHKVPEQAL"

BASE COUNT 503 a 530 C 439 G 441 T  
ORIGIN

Query Match 4.8%; Score 21; DB 48; Length 1913;  
Best Local Similarity 83.9%; Pred. No. 5.68e+00;  
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 657 atagtcgtggctgggtcatctcctccatcc 687  
||| ||||| ||||| ||||| |||||  
Cp 430 ATATCTGCGCTGGGTCTTCTCCTGCATCC 400

Search completed: Tue Dec 2 17:21:44 1997  
Job time : 415 secs.



\*\*\*\*\*

W P E R E H

(TM)

\*\*\*\*\*

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

MPSrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 2 17:22:01 1997; MasPar time 56.86 Seconds  
803.956 Million cell updates/sec

Molecular output not generated.

Title: >US-08-915-004-14  
Description: (1-438) from US08915004.seq  
Perfect Score: 438  
N.A. Sequence: 1 ATGACAAAGTGTGCTGCTG.....AGCCACAGATATGATCTGA 438  
Comp: TACTGTGTCACGACGACGAC.....TCGGTGCTATACATAGACT

Scoring table: TABLE default  
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 142080 seqs, 52183452 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: n-geneseq28  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29

Statistics: Mean 8.240; Variance 4.524; scale 1.822

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	438	100.0	438	28	T36689 Osteoclastogenesis in	0.00e+00
2	402	91.8	1173	28	T35475 Human tumour necrosis	4.15e-284
3	400	91.3	432	27	T33176 Mutated OCIF, OCIF-CC	1.48e-282
4	400	91.3	564	27	T33180 Mutated OCIF, OCIF-CB	1.48e-282
5	400	91.3	594	27	T33175 Mutated OCIF, OCIF-CD	1.48e-282
6	400	91.3	819	27	T33174 Mutated OCIF, OCIF-CD	1.48e-282
7	400	91.3	966	27	T33179 Mutated OCIF, OCIF-CC	1.48e-282
8	400	91.3	981	27	T33170 Mutated OCIF, OCIF-DD	1.48e-282
9	400	91.3	1056	27	T33173 Mutated OCIF, OCIF-DD	1.48e-282
10	400	91.3	1056	27	T33173 Mutated OCIF, OCIF-CC	1.48e-282
11	400	91.3	1080	27	T33169 Mutated OCIF, OCIF-DC	1.48e-282
12	400	91.3	1182	27	T33178 Mutated OCIF, OCIF-CB	1.48e-282
13	400	91.3	1200	27	T33172 Mutated OCIF, OCIF-CL	1.48e-282
14	400	91.3	1206	27	T33164 Mutated OCIF, OCIF-C2	1.48e-282
15	400	91.3	1206	28	T36685 Osteoclastogenesis in	1.48e-282

16	399	91.1	465	28	T36688	Osteoclastogenesis in	8.83e-282
17	398	90.9	1206	28	T33162	Mutated OCIF, OCIF-C2	5.27e-281
18	398	90.9	1206	27	T33163	Mutated OCIF, OCIF-C1	5.27e-281
19	398	90.9	1206	28	T33161	Mutated OCIF, OCIF-C2	5.27e-281
20	398	90.9	1206	27	T33165	Mutated OCIF, OCIF-C2	5.27e-281
21	372	84.9	10190	27	T33183	Fragment of human OCI	7.59e-261
22	318	72.6	321	27	T33177	Mutated OCIF, OCIF-CC	4.44e-216
23	314	71.7	1080	27	T33168	Mutated OCIF, OCIF-DC	5.43e-219
24	244	55.7	255	27	T33181	Mutated OCIF, OCIF-CP	3.81e-162
25	213	48.6	1083	27	T33166	Mutated OCIF, OCIF-DC	1.91e-138
26	187	42.7	1080	27	T33167	Mutated OCIF, OCIF-DC	1.15e-118
27	46	10.5	1047	2	Q10572	Human Natriuretic Pep	1.36e-15
28	46	10.5	1047	2	Q10572	Human Natriuretic Pep	1.36e-15
29	42	9.6	204	1	N81164	Base substituted E.co	5.65e-13
30	40	9.1	91	9	Q51746	Oligonucleotide probe	1.10e-11
31	37	8.4	91	9	Q51746	Oligonucleotide probe	8.90e-10
32	37	8.4	204	1	N81164	Base substituted E.co	8.90e-10
33	36	8.2	114	12	Q70470	Generic DNA sequence	3.77e-09
34	36	8.2	114	12	Q70469	Generic DNA sequence	3.77e-09
35	36	8.2	114	12	Q70465	Generic DNA sequence	3.77e-09
36	35	8.0	114	12	Q70467	Generic DNA sequence	1.58e-08
37	34	7.8	114	12	Q70467	Generic DNA sequence	6.57e-08
38	33	7.5	114	12	Q70468	Generic DNA sequence	2.70e-07
39	33	7.5	114	12	Q70466	Generic DNA sequence	2.70e-07
40	33	7.5	114	12	Q70468	Generic DNA sequence	2.70e-07
41	32	7.3	114	12	Q70473	Generic DNA sequence	1.09e-06
42	32	7.3	114	12	Q70469	Generic DNA sequence	1.09e-06
43	32	7.3	114	12	Q70465	Generic DNA sequence	1.09e-06
44	31	7.1	114	12	Q70472	Generic DNA sequence	4.37e-06
45	30	6.8	114	12	Q70471	Generic DNA sequence	1.73e-05

## ALIGNMENTS

RESULT 1  
ID T36689 standard; DNA; 438 BP.  
AC T36689;  
DT 22-APR-1997 (first entry)  
DE Osteoclastogenesis inhibitory factor 5 coding sequence.  
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT sig\_peptide 1..63  
FT /\*tag=a  
FT mat\_peptide 64..435  
FT /\*tag=b  
PD WO9626217-A1.  
PN 29-AUG-1996.  
PF 20-FEB-1996; J00374.  
PR 20-FEB-1995; JP-054977.  
PR 21-JUL-1995; JP-207508.  
PA (SNOW) SNOW BRAND MILK PROD CO LTD.  
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
DI WPI: 96-402320/40.  
PI P-PSDB: W99930.  
DT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
PT for bone resorption control, esp. treatment of osteoporosis  
PS Claim 24; Page 76; 183pp; Japanese.  
CC This sequence encodes full length osteoclastogenesis inhibitory  
CC factor (OCIF) 5. The OCIF of the invention has a molecular weight by  
CC SDS-PAGE of 60 kD under reducing conditions and 120 kD under non-  
CC reducing conditions. The protein is adsorbed onto cation-exchangers  
CC or heparin and its activity is lowered after 10 mins at 70 deg.C or  
CC 30 mins at 56 deg.C, and is lost after 10 mins at 90 deg.C. OCIF is  
CC useful in the control of bone resorption and therefore in the  
CC treatment and prevention of disorders of bone resorption, e.g.  
CC osteoporosis.  
SQ Sequence 438 BP; 117 A; 117 C; 110 G; 94 T;  
Query Match 100.0%; Score 438; DB 28; Length 438;  
Best Local Similarity 100.0%; Pred. No. 0.00e+00;



DR p-PSDB; R99945.  
PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
PT for bone resorption control, esp. treatment of osteoporosis  
PS Cialm 69; Page -; 183pp; Japanese.  
CC This sequence encodes a mutated version of the full length  
CC osteoclastogenesis inhibitory factor (OCIF) of the invention. The  
CC sequence encodes OCIF-CD1 in which amino acids 177-380 of the mat-  
CC protein have been deleted. The OCIF of the invention has a molecu-  
CC weight by SDS-PAGE of 60 kD under reducing conditions and 120 kD u-  
CC non-reducing conditions. The protein is adsorbed onto cation-exch-  
CC or heparin and its activity is lowered after 10 mins at 70 deg.C or  
CC mins at 56 deg.C, and is lost after 10 mins at 90 deg.C. OCIF is  
CC in the control of bone resorption and therefore in the treatment a-  
CC prevention of disorders of bone resorption, e.g. osteoporosis. The  
CC sequence is not given in the specification and is derived from the





RESULT	8	
ID	T33170 standard; DNA; 981 BP.	
AC	T33170;	
DE	22-APR-1997 (first entry)	
DE	Mutated OCIF, OCIF-DDD1, coding sequence.	
KW	Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;	
OS	osteoporosis; ss.	
KW	Synthetic.	
OS		
FH	Key	Location/Qualifiers
FT	sig_peptide	1..63
FT	/*tag= a	
FT	mat_peptide	64..978
FT	/*tag= b	
	/*product= OCIF-DDD1	
	WO9626217-Al.	
	29-AUG-1996.	
PF	20-FEB-1996; J00374.	
PR	20-FEB-1996; JP-054977.	
PR	21-JUL-1995; JP-207508.	
PA	(SNOW) SNOW BRAND MILK PROD CO LTD.	
PI	Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;	
PI	Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;	
DR	WPI: 96-402320/40.	
DR	P-PSDB; R99940.	
PT	DNA encoding osteoclastogenesis inhibitory factor protein - useful	
PT	for bone resorption control, esp. treatment of osteoporosis	
PS	Claim 54; Page 141-142; 183pp; Japanese.	
CC	This sequence encodes a mutated version of the full length	
CC	osteoclastogenesis inhibitory factor (OCIF) of the invention. This	
CC	sequence encodes OCIF-DDD1 in which amino acids 178-252 of the mature	
CC	protein have been deleted. The OCIF of the invention has a molecular	
CC	weight by SDS-PAGE of 60 kD under reducing conditions and 120 kD under	
CC	non-reducing conditions. The protein is adsorbed onto cation-exchangers	
CC	or heparin and its activity is lowered after 10 mins at 70 deg.C or 30	
CC	mins at 56 deg.C, and is lost after 10 mins at 90 deg.C. OCIF is useful	
CC	in the control of bone resorption and therefore in the treatment and	
CC	prevention of disorders of bone resorption, e.g. osteoporosis.	

ID	T3171 standard: DNA; 984 BP.
AC	T3171:
DT	22-APR-1997 (first entry)
DE	Mutated OCIF, OCIF-DDB2, coding sequence.
KW	Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;
KW	osteoporosis; ss.
OS	Synthetic.
FF	Key
FT	Location/Qualifiers
FT	sig_peptide 1..63
FT	*tag= a
FT	mat_peptide 64..981
FT	*tag= b
FT	/product= OCIF-DDD2
FT	/sig_peptide 1..63
PN	WO9626217-A1.
PD	29-AUG-1996.
PP	20-FEB-1996; J00374.
PR	20-FEB-1995; JP-054977.
PR	21-JUL-1995; JP-207508.
PR	(SNOW) SNOW BRAND MILK PROD CO LTD.
PA	Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;
PI	Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;
PI	WPI: 96-402320/40.
DR	P-PSDB: R39941.
PT	DNA encoding osteoclastogenesis inhibitory factor protein - useful
PT	for bone resorption control, esp. treatment of osteoporosis
PS	Claim 57; Page 142-143; 183pp; Japanese.
CC	This sequence encodes a mutated version of the full length
CC	osteoclastogenesis inhibitory factor (OCIF) of the invention. This
CC	sequence encodes OCIF-DDB2 in which amino acids 253-326 of the mature
CC	protein have been deleted. The OCIF of the invention has a molecular
CC	weight by SDS-PAGE of 60 kD under reducing conditions and 120 kD under
CC	non-reducing conditions. The protein is adsorbed onto cation-exchang
CC	or heparin and its activity is lowered after 10 mins at 70 deg.C or 3
CC	mins at 56 deg.C, and is lost after 10 mins at 90 deg.C. OCIF is use
CC	in the control of bone resorption and therefore in the treatment and
CC	prevention of disorders of bone resorption, e.g. osteoporosis.
CC	Sequence 984 BP; 313 A; 236 C; 212 G; 223 T;
SO	Sequence 984 BP; 313 A; 236 C; 212 G; 223 T;



Best Local Similarity 99.8%; Pred. No. 1.48e-282;		
Matches	Conservative	0; Mismatches 1; Indels 0; Gaps 0;
DbB	1	atgaacaacttgctgctgcgcctcgctgtttcttgacatctccataagtgaaccacc 60
QY	1	ATGAACAAGTTGCTGTCTCGCGCTGCTGTTCCTGGACATCTCATTAAGTGACCACC 60
DbB	61	caggaaacgtttctccaaagtacccttcattatgacgaagaacctctcatcagctgttg 120
QY	61	CAGGAACGTTCTCCAAAGTACCTTCATTATGACGAGAAACCTCTCATCAGCTGTG 120
DbB	121	tgtgacaaatgtctctctgtgtacctaccctaaacacactgtacagcaaatggaagacc 180
QY	121	TGTGACAAATGTCTCTGTCTACTTACCTAAAAACAACACTGTACAGCAAAAGTGAAGACC 180
DbB	181	gtgtgcgcccttgcctgtcacactactacaagacagctggcacacagtgacagatgt 240
QY	181	GTGTGCGCCCTTCCCTGTACCCTACTACTACAGACAGCTGGCACACACAGTGCAGAGTGT 240
DbB	241	ctatactgcagcccgctgtcaagagagctgcagtcagtcgaagcagagatgcaatcgacc 300
QY	241	CTATACTGCAGCCCGTGTGCAGAGAGCTGCAGTACGTCAAGCAGAGTGCATATCGACC 300
DbB	301	cacaaacccgctgtgcgaatcaaggaagggcgctaccttggatagattctcgttgaaa 360
QY	301	CACAAACCGCTGTCCGAATGCAGGAAGSGCGCTACCTTGACATAGAGTTCCTGTTCAA 360
DbB	361	catagagagctgccctctgatttgagttggtgcaagtggga 402
QY	361	CATAGAGCTGCCCTCTCGATTGTGAGTGTGCAAGCTGGA 402

RESULT	ID	Location/Qualifiers
T33178	standard; DNA; 1182 BP.	
AC	T33178;	
AD	T33178;	
DT	22-APR-1997 (first entry)	
DE	Mutated OCIF, OCIF-CBst, coding sequence.	
KW	Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;	
KW	osteoporosis; ss.	
OS	Synthetic.	
FH	Key.	
FT	sig_peptide	1..63
FT	*tag= a	
FT	mat_peptide	64..1179
FT	*tag= b	
FT	/product= OCIF-CBst	
FN	W09626217-Al.	
PD	29-AUG-1996.	
PD	20-FEB-1996; J00374.	
PF	20-FEB-1995; JP-054977.	
PN	21-JUL-1995; JP-207508.	
PT	(SNOW ) SNOW BRAND MILK PROD CO LTD.	
PI	Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;	
PI	Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;	
DR	WPI: 96-402320/40.	
DR	P-PSDB; R39948.	
PT	DNA encoding osteoclastogenesis inhibitory factor protein - useful	
PT	for bone resorption control, esp. treatment of osteoporosis	
PS	Claim 78; Page 148; 183pp; Japanese.	
CC	This sequence encodes a mutated version of the full length	
CC	osteoclastogenesis inhibitory factor (OCIF) of the invention. This	
CC	sequence encodes OCIF-CBst in which Gln371 is substituted with Leu	
CC	and amino acids 373-380 of the mature protein have been deleted. The	
CC	amino acid changes have been caused by the introduction of a restriction	
CC	site. The OCIF of the invention has a molecular weight by SDS-PAGE C	
CC	60 kD under reducing conditions and 120 kD under non-reducing	
CC	conditions. The protein is adsorbed onto cation-exchangers or heparin	
CC	and its activity is lowered after 10 mins at 70 deg.C or 30 mins at 5	
CC	deg.C, and is lost after 10 mins at 90 deg.C. OCIF is useful in the	
CC	control of bone resorption and therefore in the treatment and	
CC	prevention of disorders of bone resorption, e.g. osteoporosis.	
CC	Sequence 1182 BP; 376 A; 280 C; 266 G; 260 T;	

Query Match	91.3%;	Score 400;	DB 27;	Length 1182;
Best Local Similarity	99.8%;	Pred. No. 1.48e-282;		
Matches	401;	Conservative 0;	Mismatches 1;	Indels 0; Gaps 0;
Db	1	atgaacaacctctgtgcgcgcgtcgtgtttctggacatctccatcaagtggaccacc	60	
QY	1	ATGAACAAGTTCGTGTGCTGCgcgcgtcgtgtttctggacatctccatcaagtggaccacc	60	
Db	61	cagggaacgttttctctcaaaagtaccttccattatgacgaagaaacctctccatcagctgttg	120	
QY	61	CAGGAACGTTTCTCTCAAAGTACCTTCATTATGACGAAGAAACCTCTCATCAGCTGTG	120	
Db	121	tgtgacaaatgctcctcgtgtacctacctctaaacaaacactgtcacgacaaagtgggaagacc	180	
QY	121	TGTGACAAATGTCCTCCTGCTACCTACCTTAAACAAACACTGTACAGCAAAAGTGGGAAGACC	180	
Db	181	gtgtgcgcaccttgcacctaccactactacacagacagctgcgcacaccagtgacagtggt	240	
QY	181	GTGTGCGCCCTTGCCCTGACCCACTACTACACAGACAGCTGGCACACCAGTGACGAGTGT	240	
Db	241	ctatactgcagccccgtgtgcgaagagctgcagtagctcaagcaggaagtgcgaatgcgacc	300	
QY	241	CTATACTGCAGCCCGTGTGCGAAGAGCTGCAGTAGTACGTCGAAGCAGGAGTGCATCGCACC	300	
Db	301	cacaaacgcggtgtgcgaatgcgaaggaaggcgctaccttgatgatagagttctgcgttgaaa	360	
QY	301	CACAAACCGCGTGTGCCAATGCGAAGGAAGGGCGCTACCTTGAGATAGAGTTCGTCGTGAAA	360	
Db	361	cataggagctgcacctcctcgattggattgggtgggtgcgaagctgga	402	
QY	361	CATAGGAGCTGCCCTCCTCGATTGGAGTGGTGCAGAGCTGGA	402	

RESULT	13	
ID	T33172	standard; DNA; 1200 BP.
AC	T33172;	
DT	22-APR-1997	(first entry)
DE	Mutated OCIF, OCIF-CL, coding sequence.	
KW	Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;	
OS	osteoporosis; ss.	
KS	Synthetic.	
FH	Key	Location/Qualifiers
FT	sig_peptide	1..63
FT	/*tag= a	
FT	mat_peptide	64..1197
FT	/*tag= b	
FT	/product= OCIF-CL	
PN	W09636217-A1.	
PD	28-AUG-1996.	
PF	20-FEB-1996;	J00374.
PR	20-FEB-1995;	JP-034977.
PR	21-JUL-1995;	JP-207508.
PA	(SNOW ) SNOW BRAND MILK PROD CO LTD.	
PI	Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;	
PI	Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;	
DR	WPI: 96-402320/40.	
DR	P-PSDB: R99942.	
PT	DNA encoding osteoclastogenesis inhibitory factor protein - useful	
PT	for bone resorption control, esp. treatment of osteoporosis	
PS	Claim 60; Page 143-144; 183pp; Japanese.	
CC	This sequence encodes a mutated version of the full length	
CC	osteoclastogenesis inhibitory factor (OCIF) of the invention. This	
CC	sequence encodes OCIF-CL in which amino acids 379-380 of the mature	
CC	protein have been deleted. The OCIF of the invention has a molecular	
CC	weight by SDS-PAGE of 60 kD under reducing conditions and 120 kD under	
CC	non-reducing conditions. The protein is adsorbed onto cation-exchangers	
CC	or heparin and its activity is lowered after 10 mins at 70 deg.C or 30	
CC	mins at 56 deg.C, and is lost after 10 mins at 90 deg.C. OCIF is useful	
CC	in the control of bone resorption and therefore in the treatment and	
CC	prevention of disorders of bone resorption, e.g. osteoporosis.	
CC	Sequence 1200 BP; 387 A; 283 C; 288 G; 282 T;	
SQL	Query Match	91.3%; Score 400; DB 27; Length 1200;

Best Local Similarity 99.8%; Pred. No. 1.48e-282;  
Matches 401; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 atgaacaacttctgtgtgcgcgcgtctgtttctggacatctccattaaagtgaccacc 60  
QY 1 ATGAACAAGTTGCTGTGTGCGCGCTGCGTGTCTGGACATCTCCATTAAAGTGACCACC 60  
Db 61 caggaaacgttctcccaaaagtacattatgacgagaaacacctctcatcagctgttg 120  
QY 61 CAGGAAACGTTTCCCTCCAAAGTACCTTATATGACGAGAAACCTCTCATCAGCTGTG 120  
Db 121 ttgcaaaatgtctctcctgtgtacctactaaacaaacactgtacagcaaaagtgaagacc 180  
QY 121 TGTGCAAAATGTCCTCCTGGTACCTACCTAAACAACTGTACAGCAAAAGTGAAGACC 180  
Db 181 ggtgcgccttgcctgacactactacacagacagctggcacacagtgacagtggt 240  
QY 181 GGTGCGCCCTTGCCCTGACCACTACTACAGACAGCTGGCACACAGTGACAGTG 240  
Db 241 ctatactgcagcccgctgtgcaaggagctgcagtacgtcaagcaggagtgcaatcgacc 300  
QY 241 CTATAGTGCAGCCCGCTGTGCAAGGAGCTGCAGTACGTCAAGCAGGAGTGCAATCGCACC 300  
Db 301 cacaaccgctgtgcgaatgcaaggagctgcagtacgtcaagcaggagtgcaatcgacc 360  
QY 301 CACAACCCTGTGCGAATGCAAGGAGCTGCAGTACGTCAAGCAGGAGTGCAATCGCACC 360  
Db 361 catagagctgcctcctcctgtgattggagtggtgcaagctgga 402  
QY 361 CATAGGAGTGCCTCCTCGGATTTGGAGTGTGCAAGCTGGA 402

## RESULT 14

ID T33164 standard; DNA; 1206 BP.  
AC T33164;  
DT 22-APR-1997 (first entry)  
DE Mutated OCIF, OCIF-C22S, coding sequence.  
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
OS Osteoporosis; ss.  
FH Key Location/Qualifiers  
FT sig\_peptide 1..63  
FT /\*tag= a  
FT mat\_peptide 64..1203  
FT /\*tag= b  
FT /product= OCIF-C22S  
FT WO3626217-A1.  
PD 29-AUG-1996.  
PR 20-FEB-1996; J00374.  
PR 20-FEB-1995; JP-054977.  
PR 21-JUL-1995; JP-207508.  
PA (SNOW) SNOW BRAND MILK PROD CO LTD.  
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
DR WPI: 96-402320/40.  
DR P-PSDB: R99934.  
PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
PT for bone resorption control, esp. treatment of osteoporosis  
PS Claim 36; Page 135-136; 183pp; Japanese.  
CC This sequence encodes a mutated version of the full length  
CC osteoclastogenesis inhibitory factor (OCIF) of the invention. This  
CC sequence encodes OCIF-C22S in which the 22nd Cys residue in the mature  
CC OCIF protein is substituted by Ser. The OCIF of the invention has a  
CC molecular weight by SDS-PAGE of 60 kD under reducing conditions and  
CC 120 kD under non-reducing conditions. The protein is adsorbed onto  
CC cation-exchangers or heparin and its activity is lowered after 10 mins  
CC at 70 deg.C or 30 mins at 56 deg.C, and is lost after 10 mins at 90  
CC deg.C. OCIF is useful in the control of bone resorption and therefore  
CC in the treatment and prevention of disorders of bone resorption, e.g.  
CC osteoporosis.  
SQ Sequence 1206 BP; 389 A; 285 C; 268 G; 264 T;

Query Match 91.3%; Score 400; DB 27; Length 1206;  
Best Local Similarity 99.8%; Pred. No. 1.48e-282;  
Matches 401; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Best Local Similarity 99.8%; Pred. No. 1.48e-282;  
Matches 401; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 atgaacaacttctgtgtgcgcgcgtctgtttctggacatctccattaaagtgaccacc 60  
QY 1 ATGAACAAGTTGCTGTGTGCGCGCTGCGTGTCTGGACATCTCCATTAAAGTGACCACC 60  
Db 61 caggaaacgttctcccaaaagtacattatgacgagaaacacctctcatcagctgttg 120  
QY 61 CAGGAAACGTTTCCCTCCAAAGTACCTTATATGACGAGAAACCTCTCATCAGCTGTG 120  
Db 121 ttgcaaaatgtctctcctgtgtacctactaaacaaacactgtacagcaaaagtgaagacc 180  
QY 121 TGTGCAAAATGTCCTCCTGGTACCTACCTAAACAACTGTACAGCAAAAGTGAAGACC 180  
Db 181 ggtgcgccttgcctgacactactacacagacagctggcacacagtgacagtggt 240  
QY 181 GGTGCGCCCTTGCCCTGACCACTACTACAGACAGCTGGCACACAGTGACAGTG 240  
Db 241 ctatactgcagcccgctgtgcaaggagctgcagtacgtcaagcaggagtgcaatcgacc 300  
QY 241 CTATAGTGCAGCCCGCTGTGCAAGGAGCTGCAGTACGTCAAGCAGGAGTGCAATCGCACC 300  
Db 301 cacaaccgctgtgcgaatgcaaggagctgcagtacgtcaagcaggagtgcaatcgacc 360  
QY 301 CACAACCCTGTGCGAATGCAAGGAGCTGCAGTACGTCAAGCAGGAGTGCAATCGCACC 360  
Db 361 catagagctgcctcctcctgtgattggagtggtgcaagctgga 402  
QY 361 CATAGGAGTGCCTCCTCGGATTTGGAGTGTGCAAGCTGGA 402

## RESULT 15

ID T36685 standard; DNA; 1206 BP.  
AC T36685;  
DT 22-APR-1997 (first entry)  
DE Osteoclastogenesis inhibitory factor coding sequence.  
KW Osteoclastogenesis inhibitory factor; OCIF; heparin; bone resorption;  
OS Osteoporosis; ss.  
FH Key Location/Qualifiers  
FT sig\_peptide 1..63  
FT /\*tag= a  
FT mat\_peptide 64..1203  
FT /\*tag= b  
FT /label= Claim 6  
FT WO9626217-A1.  
PD 29-AUG-1996.  
PR 20-FEB-1996; J00374.  
PR 20-FEB-1995; JP-054977.  
PR 21-JUL-1995; JP-207508.  
PA (SNOW) SNOW BRAND MILK PROD CO LTD.  
PI Goto M, Higashio K, Kobayashi F, Mochizuki S, Morinaga T;  
PI Nakagawa N, Shima N, Tsuda E, Ueda M, Yano K, Yasuda H;  
DR WPI: 96-402320/40.  
DR P-PSDB: R99924-25.  
PT DNA encoding osteoclastogenesis inhibitory factor protein - useful  
PT for bone resorption control, esp. treatment of osteoporosis  
PS Claim 8; Page 66-67; 183pp; Japanese.  
CC This sequence encodes the full length osteoclastogenesis inhibitory  
CC factor (OCIF) of the invention. The OCIF has a molecular weight by  
CC SDS-PAGE of 60 kD under reducing conditions and 120 kD under non-  
CC or heparin and its activity is lowered after 10 mins at 70 deg.C or  
CC 30 mins at 56 deg.C, and is lost after 10 mins at 90 deg.C. OCIF is  
CC useful in the control of bone resorption and therefore in the  
CC treatment and prevention of disorders of bone resorption, e.g.  
CC osteoporosis.  
SQ Sequence 1206 BP; 388 A; 284 C; 269 G; 265 T;

Query Match 91.3%; Score 400; DB 28; Length 1206;  
Best Local Similarity 99.8%; Pred. No. 1.48e-282;  
Matches 401; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 atgaacaacttgctgtgcgcgtcgtgttcttgacatctccattaaagtggaccacc 60  
QY 1 ATGAACAAGTTGCTGTGTCGGCGTCTGTGTTCTGGACATCTCCATTAAATGGACACC 60  
Db 61 caggaacggtttccctccaaagtacaccttattatgacgaagaacacctctcatcagctgttg 120  
QY 61 CAGGAACGTTTCCCTCCAAAGTACCTTCATTATGACGAAGAAGCTCTCATCAGCTGTTG 120  
Db 121 tctgacaaatgtctctcctgggtacctacctaataaacaacactgtacagcaaatggagacc 180  
QY 121 TGTGACAAATGTCTCTCCCTGGGTACCTTACCTAAACAACACTGTACAGCAAAAGTGAAGACC 180  
Db 181 gftggtgcccccttgccctgaccactactacacagacagctggcacaccagtgacgaggtg 240  
QY 181 GTGTGGGCCCCCTTGCCCTGACCACCTACTACACAGACAGCTGGCACACCAGTGACGAGTGT 240  
Db 241 ctatactgagccccctgtgtgaagagctgcagtcacgtcaagcagggagtgcaatcgacc 300  
QY 241 CTATACTGAGCCCCCTGTGTGAAGAGCTGCAGTACGTCAAGCAGGAGTGCAATCGCACC 300  
Db 301 cacaaccggtgtgcgaatgcaagaaggcgctaccttgagatagattctctgttga 360  
QY 301 CACAACCGGTGTGCGAATGCAAGAAGGCGCTACCTTGAGATAGATTCTCTGTGTGAAA 360  
Db 361 cataggagctgccctcctggatttgagtggtgcaagctgga 402  
QY 361 CATAGGAGCTGCCCTCCTGGATTGGAGTGGTGGCAAGCTGGA 402

Search completed: Tue Dec 2 17:23:02 1997  
Job time : 61 secs.



\*\*\*\*\*  
M P S R C H (TM)  
\*\*\*\*\*

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by Intelligent Genetics, Inc.

MPSrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm  
Run on: Tue Dec 2 17:23:26 1997; MasPar time 278.32 Seconds  
807.828 Million cell updates/sec  
ular output not generated.

File:  
>US-08-915-004-14  
(1-438) from US08915004.seq  
Perfect Score:  
N.A. Sequence:  
Comp: 1 ATGAACAAGTGTCTGTCTG.....AGCCACAGATATCTACTCA 438  
TACTGTGTACACGACGAC.....TCGGTGTCTATACAGACT

Scoring table:  
Gap 6  
TABLE default

Nmatch STD : Dbase 0; Query 0  
707517 seqs, 256659330 bases x 2

Searched:  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database:  
EST-STS  
1. EST10 2. EST2 3. EST3 4. EST4 5. EST5 6. EST6 7. EST7 8. EST8  
9. EST9 10. EST10 11. EST11 12. EST12 13. EST13 14. EST14  
15. EST15 16. EST16 17. EST17 18. EST18 19. EST19 20. EST20  
21. EST21 22. EST22 23. EST23 24. EST24 25. EST25 26. EST26  
27. EST27 28. EST28 29. EST29 30. EST30 31. EST31 32. EST32  
33. EST33 34. EST34 35. EST35 36. EST36 37. EST37 38. EST38  
39. EST39 40. EST40 41. EST41 42. EST42 43. EST43 44. EST44  
45. EST45 46. EST46 47. EST47 48. EST48 49. EST49 50. EST50  
51. EST51 52. EST52 53. EST53 54. EST54 55. EST55 56. EST56  
57. EST57 58. EST58 59. EST59 60. EST60 61. EST61 62. EST62  
63. EST63 64. EST64 65. EST65 66. EST66 67. EST67 68. EST68  
69. EST69 70. EST70 71. EST71 72. EST72 73. EST73 74. EST74  
75. EST75 76. EST76 77. EST77 78. EST78 79. EST79 80. EST80  
81. EST81 82. EST82 83. EST83 84. EST84 85. EST85 86. EST86  
87. EST87 88. EST88 89. EST89 90. EST90 91. EST91 92. EST92  
93. EST93 94. EST94 95. EST95 96. EST96 97. EST97 98. EST98  
99. EST99  
EST-STS-TWO  
100. EST100 101. EST101 102. EST102 103. EST103 104. EST104  
105. EST105 106. EST106 107. EST107 108. EST108 109. EST109  
110. EST110 111. EST111 112. EST112 113. EST113 114. EST114  
115. EST115 116. EST116 117. EST117 118. EST118 119. EST119  
120. EST120 121. EST121 122. EST122 123. EST123 124. EST124  
125. EST125 126. EST126 127. EST127 128. EST128 129. EST129  
130. EST130 131. EST131 132. EST132 133. EST133 134. EST134  
135. EST135 136. EST136 137. EST137 138. EST138 139. EST139  
140. EST140 141. EST141 142. EST142 143. EST143 144. EST144  
145. EST145 146. EST146 147. EST147 148. EST148 149. EST149  
150. EST150 151. EST151 152. EST152 153. EST153 154. EST154  
155. EST155 156. EST156 157. EST157 158. EST158 159. EST159  
160. EST160 161. EST161 162. EST162 163. EST163 164. EST164  
165. EST165 166. EST166 167. EST167 168. EST168 169. EST169  
170. EST170 171. EST171 172. EST172 173. EST173 174. EST174

Database:  
EST-STS-TWO  
100. EST100 101. EST101 102. EST102 103. EST103 104. EST104  
105. EST105 106. EST106 107. EST107 108. EST108 109. EST109  
110. EST110 111. EST111 112. EST112 113. EST113 114. EST114  
115. EST115 116. EST116 117. EST117 118. EST118 119. EST119  
120. EST120 121. EST121 122. EST122 123. EST123 124. EST124  
125. EST125 126. EST126 127. EST127 128. EST128 129. EST129  
130. EST130 131. EST131 132. EST132 133. EST133 134. EST134  
135. EST135 136. EST136 137. EST137 138. EST138 139. EST139  
140. EST140 141. EST141 142. EST142 143. EST143 144. EST144  
145. EST145 146. EST146 147. EST147 148. EST148 149. EST149  
150. EST150 151. EST151 152. EST152 153. EST153 154. EST154  
155. EST155 156. EST156 157. EST157 158. EST158 159. EST159  
160. EST160 161. EST161 162. EST162 163. EST163 164. EST164  
165. EST165 166. EST166 167. EST167 168. EST168 169. EST169  
170. EST170 171. EST171 172. EST172 173. EST173 174. EST174

Statistics: Mean 10.031; Variance 1.782; scale 5.629

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES									
Result No.	Score	Query %		DB	ID	Description	Pred. No.		
		Match	Length						
1	23	5.3	344	61	H14106	ym62a05.r1 Homo sapie	2.49e-07		
2	22	5.0	300	183	A100384	zn46h08.r1 Stratagene	6.10e-06		
c 3	22	5.0	453	111	N21157	yx47d01.s1 Homo sapie	6.10e-06		
	22	5.0	472	57	T42477	5740 Arabidopsis thal	6.10e-06		
5	22	5.0	493	55	R1C815559A	Rice cDNA, partial se	6.10e-06		
c 6	22	5.0	687	137	AA044741	zk67g10.s1 Soares pre	6.10e-06		
7	22	5.0	796	117	W29026	55c10 Human retina CD	6.10e-06		
c 8	21	4.8	317	77	CELK094GZR	C.elegans cDNA clone	1.34e-04		
c 9	21	4.8	325	19	T54964	yb42d03.r1 Homo sapie	1.34e-04		
10	21	4.8	328	168	W40169	zc82h06.r1 Pancreat	1.34e-04		
11	21	4.8	360	166	T23370	5c08h02 membrane-free	1.34e-04		
c 12	21	4.8	371	99	N61165	TgESTzy27b03.r1 Toxop	1.34e-04		
c 13	21	4.8	376	5	T71088	yc50d04.r1 Homo sapie	1.34e-04		
c 14	21	4.8	383	47	H45707	yp23h05.s1 Homo sapie	1.34e-04		
c 15	21	4.8	395	5	T71079	yc50c04.r1 Homo sapie	1.34e-04		
16	21	4.8	441	26	R53127	y986a05.r1 Homo sapie	1.34e-04		
17	21	4.8	447	184	AA011021	ze34c01.s1 Soares ret	1.34e-04		
18	21	4.8	447	156	AA021559	ze9b06.s1 Soares ret	1.34e-04		
c 19	21	4.8	491	183	AA098806	zn44d05.s1 Stratagene	1.34e-04		
20	21	4.8	578	137	AA044796	zk67g10.r1 Soares pre	1.34e-04		
21	20	4.6	228	50	R83564	yq12e06.r1 Homo sapie	2.59e-03		
22	20	4.6	271	4	T69650	y44c09.r1 Homo sapie	2.59e-03		
c 23	20	4.6	275	107	HSC2WC112	H. sapiens partial CD	2.59e-03		
24	20	4.6	323	5	T71171	yc52c03.r1 Homo sapie	2.59e-03		
25	20	4.6	331	2	T60563	yb91c08.r1 Homo sapie	2.59e-03		
26	20	4.6	349	5	T27764	yc51a10.r1 Homo sapie	2.59e-03		
27	20	4.6	369	71	H83210	yq47d03.r1 Homo sapie	2.59e-03		
c 28	20	4.6	374	78	R98014	yq73f02.s1 Homo sapie	2.59e-03		
c 29	20	4.6	375	146	W06794	SMES0430 Schistosoma	2.59e-03		
c 30	20	4.6	377	177	AA067387	26232 Lambda-PRU2 Ara	2.59e-03		
c 31	20	4.6	390	174	W79653	zd73h08.r1 Soares fet	2.59e-03		
c 32	20	4.6	411	174	W79538	zd81g01.s1 Soares fet	2.59e-03		
c 33	20	4.6	411	17	T48207	yb44a06.s1 Homo sapie	2.59e-03		
34	20	4.6	412	185	AA135635	zl24d10.s1 Soares pre	2.59e-03		
c 35	20	4.6	420	197	H94319	yv18b09.s1 Soares fet	2.59e-03		
c 36	20	4.6	429	75	H97994	yw06e04.s1 Homo sapie	2.59e-03		
c 37	20	4.6	429	53	R92022	yp96e05.s1 Homo sapie	2.59e-03		
c 38	20	4.6	435	68	H72171	yf99g06.r1 Homo sapie	2.59e-03		
39	20	4.6	441	165	C20472	Rice cDNA, partial se	2.59e-03		
40	20	4.6	444	165	C20089	Rice cDNA, partial se	2.59e-03		
c 41	20	4.6	453	104	N74503	yz44h03.s1 Homo sapie	2.59e-03		
c 42	20	4.6	457	74	H93069	yv08h02.s1 Homo sapie	2.59e-03		
c 43	20	4.6	460	193	AA166259	ms49d11.r1 Life Tech	2.59e-03		
c 44	20	4.6	472	83	H50943	yp84a03.r1 Homo sapie	2.59e-03		
c 45	20	4.6	614	101	N63920	yz28b09.s1 Homo sapie	2.59e-03		

ALIGNMENTS

RESULT 1 H14106 344 bp mRNA EST 10-JUL-1995

LOCUS ym62a05.r1 Homo sapiens cDNA clone 163472 5' similar to SP.S32367

DEFINITION S32367 ALFA-SNAP PROTEIN - ;

ACCESSION H14106

NID 9878954

KEYWORDS EST.







## COMMENT

Contact: Wilson RK  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -40M13 fwd. from Amersham  
High quality sequence stop: 477.

## FEATURES

source

1..687  
/organism="Homo sapiens"  
/note="Organ: uterus; Vector: pT7T3-Pac; Site\_1: Not I;  
Site\_2: Eco RI; 1st strand cDNA was primed with a Not I -  
oligo (dT) primer [5',  
AACTGGAAGATTCCGGCCGCCCTTTTCTTTTCTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT7T3 vector. Library  
went through one round of normalization. Library  
constructed by M. Fatima Bonaldo."  
/clone="487938"  
/clone\_lib="Soares pregnant uterus NBHPU"  
/sex="female"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
complement(<1..>687)

## mrna

BASE COUNT

ORIGIN

179 a 162 c 160 g 178 t 8 others  
Query Match 5.0%; Score 22; DB 137; Length 687;  
Best Local Similarity 76.5%; Pred. No. 6.10e-06;  
Matches 39; Conservative 0; Mismatches 11; Indels 1; Gaps 1;  
Db 186 atacatatcagctctgtttgtcattcttcagtgagcttcccaactcc 236  
Cp 434 ATACATACTGTG-GCTTGGGTCCTCTCGATCCAGCTTGCACCTCC 385

## RESULT

LOCUS

DEFINITION

ACCESSION

NID

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

JOURNAL

COMMENT

W29026 796 bp mRNA EST 08-MAY-1996  
55c10 Human retina cDNA randomly primed sublibrary Homo sapiens  
cDNA.  
W29026  
g1308983  
EST.  
SOURCE  
human.  
Homo sapiens  
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;  
Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 796)  
Macke, J., Smallwood, P. and Nathans, J.  
Adult Human Retina cDNA  
Unpublished (1996)

## FEATURES

source

Contact: Dr. Jeremy Nathans  
Dr. Jeremy Nathans, Dept. of Molecular Biology and Genetics  
Johns Hopkins School of Medicine  
725 North Wolfe Street, Baltimore, MD 21205  
Tel: 410 955 4678  
Fax: 410 614 0827  
Email: jeremy.nathans@jhu.edu  
Clones from this library are NOT available.  
PCR Primers  
FORWARD: CTTTGGCAAGTTGACCTGGTTAAGT  
BACKWARD: GAGTGGCTTATGATGATTCTTCAGGGTAA  
Seq primer: GGTTAAAGCAAGAAGT.  
Location/Qualifiers  
1..796  
/organism="Homo sapiens"

## FEATURES

source

Query Match 4.8%; Score 21; DB 77; Length 317;  
Best Local Similarity 70.8%; Pred. No. 1.34e-04;  
Matches 34; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Db 147 tctctgcttttgcattctcactctcatccacacggnaagctct 194  
Cp 411 TCTCTGATTCAGCTTGCACCACTCCAAATCCAGGAGGCAGCTCT 364

/note="Organ: eye; Vector: lambda gt10; Site\_1: EcoRI;  
Site\_2: EcoRI; The library used for sequencing was a  
sublibrary derived from a human retina cDNA library.  
Inserts from retina cDNA library DNA were isolated,  
randomly primed, PCR amplified, size-selected, and cloned  
into lambda gt10. Individual plaques were arrayed and  
used as templates for PCR amplification, and these PCR  
products were used for sequencing."  
/clone\_lib="Human retina cDNA randomly primed sublibrary"  
/sex="mixed (males and females)"  
/tissue\_type="retina"  
/dev\_stage="adult"  
/lab\_host="E. coli strain K802"  
1..796

BASE COUNT 228 a 181 c 183 g 148 t 56 others  
ORIGIN

Query Match 5.0%; Score 22; DB 117; Length 796;  
Best Local Similarity 73.9%; Pred. No. 6.10e-06;  
Matches 34; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Db 92 gcaaggaatgcagctatttatgcagctgagctgtaccacaa 137  
QY 260 GCAAGGAGCTGCAGTACGTCAAGCAGAGTGCAATCGCACCAAA 305

## RESULT

LOCUS

DEFINITION

ACCESSION

NID

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

JOURNAL

COMMENT

8 CEUK094GZR 317 bp mRNA EST 11-DEC-1995  
C.elegans cDNA clone yk94g12 : 3' end, single read.  
D66283  
g1116674  
EST(expressed sequence tag).  
Caenorhabditis elegans (strain CB1489 him-8(el489), ) (library:  
Yuji Kohara unpublished cDNA) Hermaphrodite, male varied whole  
animal cDNA to mRNA.  
Caenorhabditis elegans  
Eukaryotes; mitochondrial eukaryotes; Metazoa; Nematoda;  
Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea;  
Rhabditidae; Caenorhabditis.  
1 (bases 1 to 317)  
Kohara, Y., Motoshima, T., Tabara, H., Sugimoto, A., Watanabe, H. and  
Nishigaki, A.  
Toward an expression map of the C.elegans genome  
Unpublished (1995)  
Submitted (23-Aug-1995) to DDBJ by:  
Yuji Kohara  
Gene Library Lab.  
National Institute of Genetics  
Yata 1111, Mishima Shizuoka  
411 Japan  
Phone: 0559-81-6854  
Fax: 0559-81-6855  
Email: ykohara@dbj.nig.ac.jp.  
Location/Qualifiers  
1..317  
/organism="Caenorhabditis elegans"  
/strain="CB1489 him-8(el489)"  
/dev\_stage="varied"  
/sex="Hermaphrodite, male"  
/tissue\_type="whole animal"  
/clone\_lib="Yuji Kohara unpublished cDNA"  
79 a 82 c 64 g 85 t 7 others

BASE COUNT 79 a 82 c 64 g 85 t 7 others  
ORIGIN

Query Match 4.8%; Score 21; DB 77; Length 317;  
Best Local Similarity 70.8%; Pred. No. 1.34e-04;  
Matches 34; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Db 147 tctctgcttttgcattctcactctcatccacacggnaagctct 194  
Cp 411 TCTCTGATTCAGCTTGCACCACTCCAAATCCAGGAGGCAGCTCT 364

RESULT 9  
 LOCUS T54964 325 bp mRNA EST 08-FEB-1995  
 DEFINITION yb42d03.r1 Homo sapiens cDNA clone 73829 5' similar to  
 gb:J03910\_rnal human (HUMAN).  
 ACCESSION T54964  
 NID G656825  
 KEYWORDS EST.  
 SOURCE human clone=73829 library-Stratagene fetal spleen (#937205)  
 vector-pBluescript SK- host-SOLR cells (kanamycin resistant)  
 primer-M13RP1 Rsttel-EcoRI Rsttel-XhoI Pooled fetal spleens. Cloned  
 unidirectionally. Primer: Oligo dt. Average insert size: 1.0 kb;  
 Uni-ZAP XR Vector; 5' adaptor sequence: 5'-GAATTCGGCAGCAG-3'; 3'  
 adaptor sequence: 5'-CTCGAGTGTGTGTGTGTGTGTGTGTGT-3'.  
 ORGANISM Homo sapiens  
 Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 325)  
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
 Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevaskis, E.,  
 Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.  
 WashU-Merck EST Project  
 Unpublished (1995)

Contact: Wilson RK  
 WashU-Merck EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1800  
 Email: est@watson.wustl.edu  
 Source: IMAGE Consortium, LLNL  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Putative full length read.

## FEATURES

Source Location/Qualifiers  
 1..325  
 /organism="Homo sapiens"  
 /clone="73829"  
 BASE COUNT 68 a 76 c 77 g 92 t 12 others  
 ORIGIN  
 Query Match 4.8%; Score 21; DB 19; Length 325;  
 Best Local Similarity 78.4%; Pred. No. 1.34e-04;  
 Matches 29; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
 Db 50 gtgcgaatgcactctctgcaagagagctgctctct 86  
 Cp 299 gtgcgaatgcactctctgctgctgctgctctct 263

RESULT 10  
 LOCUS W40169 328 bp mRNA EST 11-OCT-1996  
 DEFINITION zc82h06.r1 Pancreatic Islet Homo sapiens cDNA clone 328859 5'.  
 ACCESSION W40169  
 NID g1324291  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryotae; Mitochondrial eukaryotes; Metazoa; Chordata;  
 Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 328)  
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
 Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevaskis, E.,  
 Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.  
 WashU-Merck EST Project  
 Unpublished (1995)

Contact: Wilson RK  
 WashU-Merck EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 INSERT Length: 1169 Std Error: 0.00  
 Seq primer: mob.REGA+ET.  
 FEATURES Location/Qualifiers  
 1..328  
 /organism="Homo sapiens"  
 /note="Organ: pancreas; Vector: pBluescript SK-; Site\_1:  
 ECORI; Site\_2: XhoI; Reference: Hum Mol Gen 2, 1795 (1993)  
 Takeda et al. Cloned unidirectionally. Primer: Oligo dt.  
 -5' adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor  
 sequence: 5' CTCGAGTGTGTGTGTGTGTGTGTGT 3'"  
 /clone="328859"  
 /clone\_lib="Pancreatic Islet"  
 /tissue\_type="pancreatic islet"  
 /lab\_host="SOLR cells (kanamycin resistant)"  
 <1..>328  
 mRNA 105 a 56 c 68 g 98 t 1 others  
 ORIGIN

Query Match 4.8%; Score 21; DB 168; Length 328;  
 Best Local Similarity 76.9%; Pred. No. 1.34e-04;  
 Matches 30; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
 Db 202 tgacctctgtattctgaatgtgaactctggaagcagg 240  
 QY 370 TGCCCTCCTGGATTGGAGTGTGCAAGCTGGAIGCAGG 408

RESULT 11  
 LOCUS T23370 360 bp mRNA EST 17-OCT-1996  
 DEFINITION 5c08h02 membrane-free polysomes from endosperm Zea mays cDNA clone  
 5c08h02 5' end.  
 ACCESSION T23370  
 NID 9511392  
 KEYWORDS EST.  
 SOURCE Zea mays.  
 ORGANISM Zea mays

Eukaryotae; Mitochondrial eukaryotes; Viridiplantae;  
 Charophyta/Embryophyta group; Embryophyta; Magnoliophyta;  
 Liliopsida; Poales; Poaceae; Zea.  
 REFERENCE 1 (bases 1 to 360)  
 AUTHORS Shen, B., Carneiro, N., Torres-Jerez, I., Stevenson, R., McCreery, T.,  
 Helentjaris, T., Baysdorfer, C., Almira, E., Ferl, R., Habben, J. and  
 Larkins, B.  
 TITLE Partial sequencing and mapping of clones from two maize cDNA  
 libraries  
 JOURNAL Plant Mol. Biol. 26, 1085-1101 (1994)  
 MEDLINE 95111093  
 COMMENT

Contact: The Maize cDNA Project  
 Helentjaris TG (primary contact)  
 Dept. of Plant Sciences  
 University of Arizona  
 Dept. of Plant Sciences, University of Arizona, Tucson, AZ, 85721  
 ph: 602-6218-746  
 fax: 602-621-7186  
 E-mail: helnjars@ccit.arizona.edu

Chris Baysdorfer  
 Department of Biological Sciences, School of Science  
 California State University, Hayward  
 Hayward, CA 94542  
 ph: 510-881-3459  
 fax: 510-727-2035  
 E-mail: cbaysdor@sl.csuhayward.edu

Rob Ferl

1

KEYWORDS  
SOURCE

EST.  
human clone-188313 library-Soares breast 3NdbHst vector-pT7n3D  
(Pharmacacia) with a modified polylinker host-DH10B (ampicillin  
resistant) primer-SP6 Rsite1-Not I Rsite2-Eco RI Adult human. 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTACCAATCTGAAGTGGGCGCGCCCTTTTTTTTTTTTTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors (Pharmacacia),  
digested with Not I and cloned into the Not I and Eco RI sites of a  
modified pT73 vector (Pharmacacia). Library went through one round  
of normalization to a Cot = 20. Library constructed by Bento Soares  
and M.Fatima Bonaldo.

## ORGANISM

Homo sapiens  
Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;  
Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;  
Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria;  
Eutheria; Archonta; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 395)  
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,  
Holman,M., Hultman,M., Kucaba,T., Le.M., Lennon,G., Marra,M.,  
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,  
Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and  
Wilson,R.

REFERENCE  
AUTHORSTITLE  
JOURNAL  
COMMENT

The WashU-Merck EST Project  
Unpublished (1995)

Contact: Wilson RK  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

High quality sequence stops: 268  
Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

## FEATURES

Source  
1..383  
/organism="Homo sapiens"  
/clone="188313"

BASE COUNT 115 a 79 c 91 g 90 t 8 others  
ORIGIN

Query Match 4.8%; Score 21; DB 47; Length 383;  
Best Local Similarity 81.8%; Pred. No. 1.34e-04;  
Matches 27; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 258 cctccacattaccgtcattaggacgtgaacc 290  
||||| | |||| | |||| | |||| | ||||  
73 CTTCCAAAGTACTCTTATTATGACGAAGAAC 105

## RESULT 15

LOCUS T71079 395 bp mRNA EST 01-MAR-1995  
DEFINITION yc50c04.r1 Homo sapiens cDNA clone 84102 5' similar to  
gb:J03910..rnl Human (HUMAN);  
T71079  
ACCESSION T71079  
NID g685600  
KEYWORDS EST.  
SOURCE human clone-84102 library-Stratagene liver (#937224)

vector-pBluescript SK host-SOLR cells (kanamycin resistant)  
Primer-M13Rpl Rsite1-ECORI Rsite2-XhoI Cloned unidirectionally.  
Primer: Oligo dT. Hepatectomy from normal 49 year old male  
caucasian. Average insert size: 1.1 kb; Uni-ZAP XR Vector; 5'  
adaptor sequence: 5'-GAATTCGCCACGAG-3'; 3' adaptor sequence:  
5'-CTCGAGTTTTTTTTTTTTTTT-3'.

## ORGANISM

Homo sapiens  
Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 395)  
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,  
Holman,M., Hultman,M., Kucaba,T., Le.M., Lennon,G., Marra,M.,  
Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,

REFERENCE  
AUTHORS

Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.  
WashU-Merck EST Project  
Unpublished (1995)

## TITLE

JOURNAL  
COMMENT

Contact: Wilson RK  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

High quality sequence stops: 304  
Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

## FEATURES

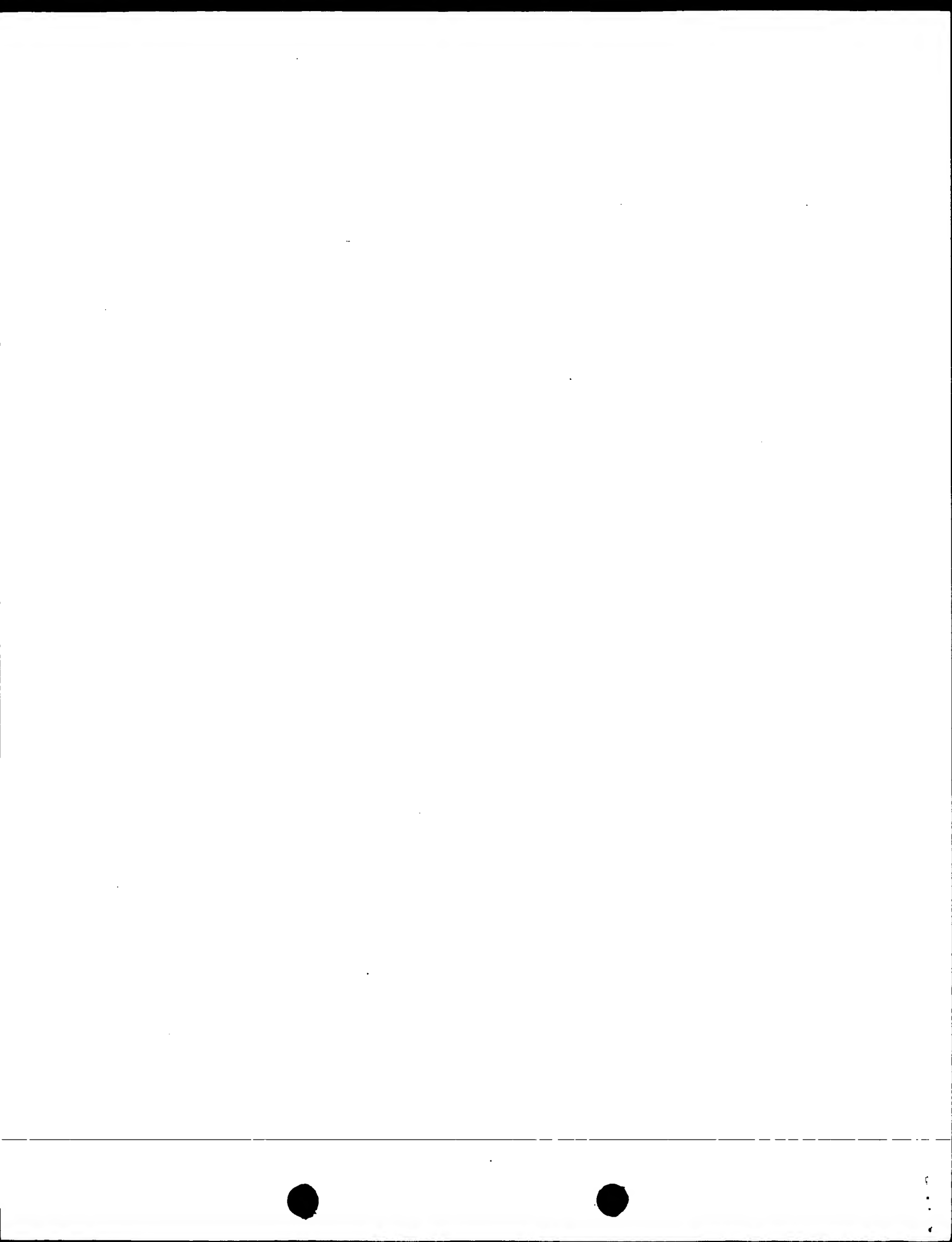
Source  
1..395  
/organism="Homo sapiens"  
/clone="84102"

BASE COUNT 85 a 105 c 91 g 110 t 4 others  
ORIGIN

Query Match 4.8%; Score 21; DB 5; Length 395;  
Best Local Similarity 78.4%; Pred. No. 1.34e-04;  
Matches 29; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Db 124 gtgcaaatgcactctctgcaagaagagctgctgctct 160  
||||| | |||| | |||| | |||| | ||||  
Cp 299 GTGCGATTGCACCTCTGCTTGCAGCTACTGCAGCTCCT 263

Search completed: Tue Dec 2 17:28:22 1997  
Job time : 296 secs.



\*\*\*\*\*

\*\*\*\*\*

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (C) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by Intelligent, Inc.

MPSrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 2 17:28:43 1997; MasPar time 144.13 seconds  
835.145 Million cell updates/sec

File: >US-08-915-004-14  
Description: (1-438) from US08915004.seq  
Perfect Score: 438  
N.A. Sequence: 1 ATGACACAGTTCGTGCTG.....AGCCACAGATATGTATCTGA 438  
Comp: TACTGTGTCACGACGAC.....TCGGTGTCTATACATAGACT

Scoring table: TABLE default  
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 359085 seqs, 137405154 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database:

EST-STS-THREE  
1:EST199 2:EST200 3:EST201 4:EST202 5:EST203 6:EST204  
7:EST205 8:EST206 9:EST207 10:EST208 11:EST209 12:EST210  
13:EST211 14:EST212 15:EST213 16:EST214 17:EST215  
18:EST216 19:EST217 20:EST218 21:EST219 22:EST220  
23:EST221 24:EST222 25:EST223 26:EST224 27:EST225  
28:EST226 29:EST227 30:EST228 31:EST229 32:EST230  
33:EST231 34:EST232 35:EST233 36:STS1 37:STS2 38:STS3  
39:STS4 40:STS5 41:STS6 42:STS7 43:STS8 44:STS9 45:STS10  
46:STS11 47:STS12 48:STS13

EST-STS-FOUR  
49:gnEST1 50:gnEST2 51:gnEST3 52:gnEST4 53:gnEST5  
54:gnEST6 55:gnEST7 56:gnEST8 57:gnEST9 58:gnEST10  
59:gnEST11 60:gnEST12 61:gnEST13 62:gnEST14 63:gnEST15  
64:gnEST16 65:gnEST17 66:gnEST18 67:gnEST19 68:gnEST20  
69:gnEST21 70:gnEST22 71:gnEST23 72:gnEST24 73:gnEST25  
74:gnEST5 75:gnEST1 76:gnEST2 77:gnEST3 78:gnEST4  
79:gnEST5 80:gnEST6 81:gnEST7 82:gnEST8 83:gnEST9  
84:gnEST10 85:gnEST11 86:gnEST12 87:gnEST13 88:gnEST14  
89:gnEST15 90:gnEST16 91:gnEST17 92:gnEST18 93:gnEST19  
94:gnEST20 95:gnEST21 96:gnEST22 97:gnEST23 98:gnEST24  
99:gnEST25 100:gnEST26 101:gnEST27 102:gnEST28  
103:gnEST29 104:gnEST30 105:gnEST31 106:gnEST32 107:ueEST1  
108:ueEST2

Statistics: Mean 10.082; Variance 1.809; scale 5.572

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result Query

\*\*\*\*\*

\*\*\*\*\*

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (C) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by Intelligent, Inc.

MPSrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 2 17:28:43 1997; MasPar time 144.13 seconds  
835.145 Million cell updates/sec

File: >US-08-915-004-14  
Description: (1-438) from US08915004.seq  
Perfect Score: 438  
N.A. Sequence: 1 ATGACACAGTTCGTGCTG.....AGCCACAGATATGTATCTGA 438  
Comp: TACTGTGTCACGACGAC.....TCGGTGTCTATACATAGACT

Scoring table: TABLE default  
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 359085 seqs, 137405154 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database:

EST-STS-THREE  
1:EST199 2:EST200 3:EST201 4:EST202 5:EST203 6:EST204  
7:EST205 8:EST206 9:EST207 10:EST208 11:EST209 12:EST210  
13:EST211 14:EST212 15:EST213 16:EST214 17:EST215  
18:EST216 19:EST217 20:EST218 21:EST219 22:EST220  
23:EST221 24:EST222 25:EST223 26:EST224 27:EST225  
28:EST226 29:EST227 30:EST228 31:EST229 32:EST230  
33:EST231 34:EST232 35:EST233 36:STS1 37:STS2 38:STS3  
39:STS4 40:STS5 41:STS6 42:STS7 43:STS8 44:STS9 45:STS10  
46:STS11 47:STS12 48:STS13

EST-STS-FOUR  
49:gnEST1 50:gnEST2 51:gnEST3 52:gnEST4 53:gnEST5  
54:gnEST6 55:gnEST7 56:gnEST8 57:gnEST9 58:gnEST10  
59:gnEST11 60:gnEST12 61:gnEST13 62:gnEST14 63:gnEST15  
64:gnEST16 65:gnEST17 66:gnEST18 67:gnEST19 68:gnEST20  
69:gnEST21 70:gnEST22 71:gnEST23 72:gnEST24 73:gnEST25  
74:gnEST5 75:gnEST1 76:gnEST2 77:gnEST3 78:gnEST4  
79:gnEST5 80:gnEST6 81:gnEST7 82:gnEST8 83:gnEST9  
84:gnEST10 85:gnEST11 86:gnEST12 87:gnEST13 88:gnEST14  
89:gnEST15 90:gnEST16 91:gnEST17 92:gnEST18 93:gnEST19  
94:gnEST20 95:gnEST21 96:gnEST22 97:gnEST23 98:gnEST24  
99:gnEST25 100:gnEST26 101:gnEST27 102:gnEST28  
103:gnEST29 104:gnEST30 105:gnEST31 106:gnEST32 107:ueEST1  
108:ueEST2

Statistics: Mean 10.082; Variance 1.809; scale 5.572

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result Query

No.	Score	Match	Length	DB	ID	Description	Pred. No.
1	22	5.0	472	77	AT4774	5740 Arabidopsis thal	4.95e-06
2	21	4.8	431	24	AA219045	zq16e09.s1 Stratagene	1.04e-04
3	20	4.6	204	106	HSC5200	Human chromosome 5 LA	1.93e-03
4	20	4.6	204	46	HUMC5200	Human chromosome 5 LA	1.93e-03
5	20	4.6	204	47	HUMC5200	Human chromosome 5 LA	1.93e-03
6	20	4.6	204	74	HUMC5200	Human chromosome 5 LA	1.93e-03
7	20	4.6	377	79	ATA67387	26232 Lambda-PRL2 Ara	1.93e-03
8	20	4.6	400	40	G14220	Human STS SHGC-7175 c	1.93e-03
9	20	4.6	417	82	HS1147657	zr55g11.s1 Soares NH	1.93e-03
10	20	4.6	417	63	AA228090	zr55g11.s1 Soares NH	1.93e-03
11	20	4.6	421	66	AA235526	zr55g11.s1 Soares ova	1.93e-03
12	20	4.6	421	83	HS1153832	zr55g11.s1 Soares ova	1.93e-03
13	20	4.6	423	3	AA112480	zn69a03.s1 Stratagene	1.93e-03
14	20	4.6	424	15	AA191137	zn69a03.s1 Stratagene	1.93e-03
15	20	4.6	503	22	AA211552	mn55f02.r1 Stratagene	1.93e-03
16	19	4.3	340	55	AA087228	mn55f06.r1 Stratagene	1.93e-03
17	19	4.3	360	25	N58571	yy55c12.r1 Soares fet	3.14e-02
18	19	4.3	362	63	AA228313	nc38f10.r1 NCI CGAP P	3.14e-02
19	19	4.3	382	82	HS1147808	nc38f10.r1 NCI CGAP P	3.14e-02
20	19	4.3	374	18	AA202904	LD03320.Sprime LD Dro	3.14e-02
21	19	4.3	388	44	G29395	human STS SHGC-32985	3.14e-02
22	19	4.3	404	18	AA201435	LD04428.Sprime LD Dro	3.14e-02
23	19	4.3	413	30	AA061856	ml33a02.r1 Stratagene	3.14e-02
24	19	4.3	416	97	MM9379	mb58a08.r1 Soares mou	3.14e-02
25	19	4.3	419	33	AA108214	ml63f11.r1 Stratagene	3.14e-02
26	19	4.3	420	50	AA138946	mr63e09.r1 Stratagene	3.14e-02
27	19	4.3	433	87	HS525332	zcl9c10.s1 Soares par	3.14e-02
28	19	4.3	444	50	AA144720	mr68c01.r1 Stratagene	3.14e-02
29	19	4.3	461	93	MM13311	mc27e09.r1 Soares mou	3.14e-02
30	19	4.3	470	26	N76192	yz30f08.r1 Soares mul	3.14e-02
31	19	4.3	473	18	AA200366	mu36c04.r1 Soares 2Nb	3.14e-02
32	19	4.3	473	99	MMAA3366	mu36c04.r1 Soares 2Nb	3.14e-02
33	19	4.3	478	32	AA097096	mm52b05.r1 Stratagene	3.14e-02
34	19	4.3	478	40	G11993	human STS DYS200	3.14e-02
35	19	4.3	487	50	AA144641	mr70d06.r1 Stratagene	3.14e-02
36	19	4.3	511	72	AA253994	va10g02.r1 Soares mou	3.14e-02
37	19	4.3	524	20	AA209457	zq82h11.r1 Stratagene	3.14e-02
38	19	4.3	526	31	AA068479	mm46b12.r1 Stratagene	3.14e-02
39	19	4.3	555	18	AA202414	LD05386.Sprime LD Dro	3.14e-02
40	19	4.3	598	69	AA246779	LD05719.Sprime LD Dro	3.14e-02
41	19	4.3	598	80	DMAA46779	LD05719.Sprime LD Dro	3.14e-02
42	19	4.3	609	88	HS995331	z8b5b07.r1 Soares sen	3.14e-02
43	19	4.3	611	22	AA212024	mo87c04.r1 Beddington	3.14e-02
44	19	4.3	646	15	AA191607	zq43e07.s1 Stratagene	3.14e-02
45	19	4.3	652	20	AA208979	mo80h01.r1 Beddington	3.14e-02

## ALIGNMENTS

RESULT 1  
ID AT4774  
AC 142477;  
NI 9933235  
DT 02-FEB-1995 (Rel. 42, Created)  
DE 12-MAR-1997 (Rel. 51, Last updated, Version 15)  
KW EST  
OC Arabidopsis thaliana (thale cress)  
OS Arabidopsis thaliana (thale cress)  
OC Eukaryotae; mitochondrial eukaryotes; Viridiplantae;  
OC Charophyta/Embryophyta group; Embryophyta; Magnoliophyta;  
OC Magnoliopsida; Capparales; Brassicaceae; Arabidopsis.  
RN [1]  
RP 1-472  
RX MEDLINE; 95148729.  
RA Newman T., deBruin F.J., Green P., Keegstra K., Kende H.,  
RA McIntosh L., Ohlrogge J., Raikhel N., Somerville S., Thomas M.,  
RA Retzel E., Somerville C.;  
RT "Genes galore: a summary of methods for accessing results from  
RT large-scale partial sequencing of anonymous Arabidopsis cDNA  
RT clones";  
RL Plant Physiol. 106:1241-1255(1994).







directionally cloned with Sal-Notarms using oligo dt prime

CDNA."  
/clone="88F5T7"  
/clone\_lib="Lambda-PRL2"  
<1..>377

Sequence 377 BP; 100 A; 93 C; 64 G; 105 T; 15 other;

Query Match 4.6%; Score 20; DB 79; Length 377;

Best Local Similarity 73.2%; Pred. No. 1.93e-03;  
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Db 177 tgcacaatgcacacacacgcgaatcgacgcacacacgc 217

QY 269 tgcagtagctcaagcaggagtgcatcgacccacacacgc 309

# RESULT

8 G14220 400 bp DNA STS 22-DEC-1995  
human STS SHGC-7175 clone pg-3149.

ACCESSION G14220

NID g1129959

STs sequence; primer; sequence tagged site.

SOURCE human.

## ORGANISM

Homo sapiens  
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;  
Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Choanata;  
Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates;  
Catarrhini; Homnidae; Homo.  
1 (bases 1 to 400)

## REFERENCE

AUTHORS Myers,R.M.

JOURNAL Unpublished (1995)

## COMMENT

Contact: Richard M. Myers  
Stanford Human Genome Center (SHGC)  
Stanford University School of Medicine  
Department of Genetics, M-344, Stanford, CA 94305, USA  
Tel: 415/7259687  
Fax: 415/7259689  
Email: myers@shgc.stanford.edu

Primer A: GGCTACACTGTGTCATAGGGA  
Primer B: GTCACTCAGGCAATTCCTACT

STs size: 94

PCR Profile:

Initial incubation: 94 degrees C for 90 seconds  
Denaturation: 94 degrees C for 15 seconds  
Annealing: 62 degrees C for 23 seconds  
Polymerization: 72 degrees C for 30 seconds  
PCR Cycles: 30  
Thermal Cycler: Perkin Elmer 9600

Protocol:

Template: 25 ng  
Primer: each 1 uM  
dNTPs: each 200 uM  
Tag Polymerase: 0.05 units/ul  
Total Vol: 10 ul

Buffer:

MgCl2: 2.5 mM

KCl: 50 mM

Tris-HCl: 20 mM

pH: 8.3

Plasmid clones, generated from a lymphoblastoid cell line from a human male. Localized to human chromosome 3 by analysis on the NIGMS Human/Rodent Somatic Cell Hybrid Panel #1, Coriell Institute for Medical Research, Camden, NJ 08103.

## FEATURES

source

1..400

/organism="Homo sapiens"

255..348

STs

22; Conservative

0; Mismatches

2; Indels

0; Gaps

0;

primer\_bind /map="3"

255..275

primer\_bind /map="3"

complement(328..348)

/map="3"

BASE COUNT 100 a 95 c 90 g 99 t 16 others

ORIGIN

Query Match 4.6%; Score 20; DB 40; Length 400;

Best Local Similarity 75.0%; Pred. No. 1.93e-03;

Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Db 177 acaccacttactccagctggaagacagagtgagaccct 216

QY 153 ACAACACTGTACGCAAACTGGAAGACCGTGTGCGCCCT 192

# RESULT

ID HS1147657 standard; RNA; EST; 417 BP.

AC AA228090;

NI g1849624

DT 27-FEB-1997 (Rel. 51, Created)

DE 27-FEB-1997 (Rel. 51, last updated, Version 1)

DE zr55g11.s1 Soares NHMPu S1 Homo sapiens cDNA clone 667364 3'

DE similar to TR:G307146 G307146 PROTEIN-LYSINE OXIDASE HOMOLOG

DE DE PRECURSOR ;

KW EST.

OS Homo sapiens (human)

OC Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;

OC Vertebrata; Euthera; Primates; Catarrhini; Homnidae; Homo.

RN [1]

RP 1-417

RA Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M.,

RA Holman M., Hultman M., Kucaba T., Le M., Lennon G., Marra M.,

RA Parsons J., Rifkin L., Rohlfing T., Soares M., Tan F.,

RA Trevaskis E., Waterston R., Williamson A., Wohlmann P., Wilson R.;

RT "The WashU-Merck EST Project";

RL Unpublished.

CC Contact: Wilton RK WashU-Merck EST Project Washington University

CC School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis,

CC MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email:

CC est@wason.wustl.edu This clone is available royalty-free through

CC LNL; contact the IMAGE Consortium (info@image.lnl.gov) for

CC further information. Possible reversed clone: similarity on wrong

CC strand seq primer: -41ml3 fwd. ET from Amersham High quality

CC sequence stop: 332.

Key

Location/Qualifiers

source

1..417

/organism="Homo sapiens"

/note="Organ: mixed (see below); Vector: pT7T3D-Pac

(Pharmacia) with a modified polylinker; Site 1: Not I;

Site 2: Eco RI; Equal amounts of plasmid DNA from three

normalized libraries (melanocyte 2NbM, pregnant uterus

NbHPU, and fetal heart NbH19W) were mixed, and ss circles

were made in vitro. Following HAP purification, this DNA

was used as tracer in a subtractive hybridization reaction

The driver was PCR-amplified cDNAs from pools of 5,000

clones made from the same 3 libraries. The pools consisted

of I.M.A.G.E. clones 260232-265223, 340488-345479, and

484488-489479.

/clone="667364"

/clone\_lib="Soares NHMPu S1"

/tissue\_type="Pooled human melanocyte, fetal heart, and

pregnant uterus"

/lab\_host="DH10B"

complement(<1..>417)

MRNA

Sequence 417 BP; 102 A; 90 C; 115 G; 109 T; 1 other;

Query Match 4.6%; Score 20; DB 82; Length 417;

Best Local Similarity 91.7%; Pred. No. 1.93e-03;

Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

Db 90 aaacaacactgtagagaaagt 113
      |||||
QY 150 AAAACAACACTGTACACAAAGT 173

RESULT 10
LOCUS AA228090 417 bp mRNA EST 24-FEB-1997
DEFINITION zr59g11.s1 Soares NhMPu S1 Homo sapiens cDNA clone 667364 3'
similar to TR:G307146 G307146 PROTEIN-LYSINE OXIDASE HOMOLOG
PRECURSOR ;
ACCESSION AA228090
NID 91849624
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 417)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)
TITLE The WashU-Merck EST Project
JOURNAL
COMMENT

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone: similarity on wrong strand
Seq primer: -41m13 fwd. ET from Amersham
High quality sequence stop: 332.
Location/Qualifiers
1..417
/organism="Homo sapiens"
/notes="Organ: mixed (see below); Vector: pT7J3D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NBHM, pregnant uterus
NhMPu, and fetal heart NhH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."
/clone="667364"
/clone_lib="Soares NhMPu S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
complement(<1..>417)
BASE COUNT 102 a 90 c 115 g 109 t 1 others
ORIGIN

Query Match 4.6%; Score 20; DB 63; Length 417;
Best Local Similarity 91.7%; Pred. No. 1.93e-03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 90 aaacaacactgtagagaaagt 113
      |||||
QY 150 AAAACAACACTGTACACAAAGT 173

RESULT 11
LOCUS AA235526 421 bp mRNA EST 03-MAR-1997
DEFINITION zr59g11.s1 Soares NhMPu S1 Homo sapiens cDNA clone 667364 3'
similar to TR:G307146 G307146 PROTEIN-LYSINE OXIDASE HOMOLOG
PRECURSOR ;
ACCESSION AA235526
NID 91859981
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 421)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)
TITLE The WashU-Merck EST Project
JOURNAL
COMMENT

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone: similarity on wrong strand
Seq primer: -41m13 fwd. ET from Amersham
High quality sequence stop: 174.
Location/Qualifiers
1..421
/organism="Homo sapiens"
/notes="Organ: ovary; Vector: pT7J3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I oligo(dT) primer [5'
TGTTACCAATCTGAGTGGGAGCGCGGTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7J3 vector
(Pharmacia). Library constructed by Bento Soares and
M.Fatima Bonaldo."
/clone="723880"
/clone_lib="Soares ovary tumor NhHOT"
/sex="female"
/tissue_type="ovarian tumor"
/lab_host="DH10B (ampicillin resistant)"
complement(<1..>421)
BASE COUNT 111 a 96 c 119 g 95 t
ORIGIN

Query Match 4.6%; Score 20; DB 66; Length 421;
Best Local Similarity 91.7%; Pred. No. 1.93e-03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 88 aaacaacactgtagagaaagt 111
      |||||
QY 150 AAAACAACACTGTACACAAAGT 173

RESULT 12
ID HS1153832 standard; RNA; EST; 421 BP.
AC AA235526;
NI 91859981
DT 06-MAR-1997 (Rel. 51, Created)
DT 06-MAR-1997 (Rel. 51, Last updated, Version 1)
DE zr59g11.s1 Soares ovary tumor NhHOT Homo sapiens cDNA clone 723880
DE 3' similar to TR:G307146 G307146 PROTEIN-LYSINE OXIDASE HOMOLOG
DE DE PRECURSOR ;
KW EST.
OS Homo sapiens (human)
OC Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;

```

OC Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP 1-421  
 RA Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M.,  
 RA Holman M., Hultman M., Kucaba T., Le M., Lennon G., Marra M.,  
 RA Parsons J., Rifkin L., Rohlfing T., Soares M., Tan F.,  
 RA Trevisakis E., Waterston R., Williamson A., Wohldmann P., Wilson R.;  
 RA "The WashU-Merck EST Project";  
 RL Unpublished.  
 CC Contact: Wilson RK WashU-Merck EST Project Washington University  
 CC School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis,  
 CC MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email:  
 CC est@watson.wustl.edu This clone is available royalty-free through  
 CC the IMAGE Consortium (info@image.llnl.gov) for  
 CC LNL; contact the IMAGE Consortium (info@image.llnl.gov) for  
 CC further information. Possible reversed clone: similarity on wrong  
 CC strand Seq primer: -41ml3 fwd. ET from Amersham High quality  
 CC sequence stop: 174.  
 CC Key  
 source 1..421  
 /organism="Homo sapiens"  
 /note="Organ: ovary; Vector: pT73D (Pharmacia) with a  
 modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer [5'  
 TGTACATCTCAAGTGGAGCGCGGCTTTTGTCTTTT 3'],  
 double-stranded cDNA was size selected, ligated to Eco RI  
 adapters (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of a modified pT73 vector  
 (Pharmacia). Library constructed by Bento Soares and  
 M. Fatima Bonaldo."  
 /clone\_lib="Soares ovary tumor NBHOT"  
 /sex="Female"  
 /tissue\_type="ovarian tumor"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /complement(<1..>421)  
 Query Match 4.6%; Score 20; DB 83; Length 421;  
 Best Local Similarity 91.7%; Pred. No. 1.93e-03;  
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Db 88 aaaaacacactgtagagaaagt 111  
 |||||  
 |||||  
 QY 150 AAAACACACTGTACAGCAAGTG 173  
 |||||  
 |||||  
 LT 13  
 LOCUS AA112480 423 bp mRNA EST 08-NOV-1996  
 DEFINITION zn69a03.s1 Stratagene HeLa cell s3 937216 Homo sapiens cDNA clone  
 563404 3'.  
 ACCESSION AA112480  
 NID G1665157  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;  
 Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 423)  
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
 Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevisakis, E.,  
 Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.  
 WashU-Merck EST Project  
 Unpublished (1995)  
 Contact: Wilson RK  
 WashU-Merck EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 TITLE  
 JOURNAL  
 COMMENT

This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -40M13 fwd. from Amersham  
 High quality sequence stop: 353.  
 FEATURES  
 source 1..423  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
 XhoI; Cloned unidirectionally. Primer: Oligo dT. HeLa S3  
 epithelioid carcinoma cells grown to semi-confluency  
 without induction. Average insert size: 1.5 kb; Uni-ZAP XR  
 vector. -5' adaptor sequence: 5' GAATTCGCGACGAG 3' -3'  
 adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'  
 /clone="563404"  
 /clone\_lib="Stratagene HeLa cell s3 937216"  
 /sex="female"  
 /dev\_stage="HeLa S3 cell line"  
 /lab\_host="SOLR (kanamycin resistant)"  
 /complement(<1..>423)  
 BASE COUNT 149 a 77 c 121 t 3 others  
 ORIGIN  
 Query Match 4.6%; Score 20; DB 3; Length 423;  
 Best Local Similarity 81.3%; Pred. No. 1.93e-03;  
 Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 Db 227 agtgaggaagcagcagctacattgtcattca 258  
 |||||  
 |||||  
 Cp 149 AGGTAGTACCAGGAGGACATTTGTCACAA 118  
 |||||  
 |||||  
 RESULT 14  
 LOCUS AA191137 424 bp mRNA EST 15-JAN-1997  
 DEFINITION zn86c11.1 Stratagene HeLa cell s3 937216 Homo sapiens cDNA clone  
 627092 5'.  
 ACCESSION AA191137  
 NID G1779831  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;  
 Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 424)  
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
 Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevisakis, E.,  
 Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.  
 WashU-Merck EST Project  
 Unpublished (1995)  
 Contact: Wilson RK  
 WashU-Merck EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -28M13 rev2 from Amersham.  
 FEATURES  
 source 1..424  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
 XhoI; Cloned unidirectionally. Primer: Oligo dT. HeLa S3  
 epithelioid carcinoma cells grown to semi-confluency  
 without induction. Average insert size: 1.5 kb; Uni-ZAP XR  
 vector. -5' adaptor sequence: 5' GAATTCGCGACGAG 3' -3'  
 adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'  
 /clone="627092"  
 /clone\_lib="Stratagene HeLa cell s3 937216"  
 /sex="female"  
 /dev\_stage="HeLa S3 cell line"

/lab\_host="SOLR (kanamycin resistant)"  
<1..>424

BASE COUNT 99 a 119 c 111 g 90 t 5 others  
ORIGIN

Query Match 4.6%; Score 20; DB 15; Length 424;  
Best Local Similarity 85.7%; Pred. No. 1.93e-03;  
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 206 caaggagctgcagtcggtggagcaggag 233  
|||||  
QY 261 CAAGGAGCTGCAGTACGTCAAGCAGGAG 288

RESULT 15 AA211552 503 bp mRNA EST 31-JAN-1997  
LOCUS zn55f02.r1 Stratagene muscle 937209 Homo sapiens cDNA clone 562107  
DEFINITION 5'

ACCESSION AA211552  
NID 91810206  
KEYWORDS EST.

ORIGIN human.

ORGANISM Homo sapiens

Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;  
Vertebrata; Euthera; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 503)  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevaskis, E.,  
Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.  
WashU-Merck EST Project  
Unpublished (1995)

TITLE  
JOURNAL  
COMMENT

Contact: Wilson RK  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: 28ml3 rev1 Et from Amersham  
High quality sequence stop: 440.  
Location/Qualifiers

FEATURES  
source

1..503  
/organism="Homo sapiens"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
XhoI; Cloned unidirectionally. Primer: Oligo dr. Skeletal  
muscle from patient with malignant hyperthermia. Average  
insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor  
sequence: 5' GAATTCGGCAGGAG 3' -3' adaptor sequence: 5'  
CTCGAGTTTTTTTTTTTTTTT 3'"

/clone="562107"  
/clone\_lib="Stratagene muscle 937209"  
/dev\_stage="adult"  
/lab\_host="SOLR (kanamycin resistant)"

BASE COUNT 123 a 136 c 135 g 104 t 5 others  
ORIGIN

Query Match 4.6%; Score 20; DB 22; Length 503;  
Best Local Similarity 85.7%; Pred. No. 1.93e-03;  
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 257 caaggagctgcagtcggtggagcaggag 284  
|||||  
QY 261 CAAGGAGCTGCAGTACGTCAAGCAGGAG 288

Search completed: Tue Dec 2 17:31:35 1997  
Job time : 172 secs.

